Plant Science 291 (2020) 110364

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Discovery of modules involved in the biosynthesis and regulation of maize phenolic compounds

Lina Gomez-Cano[a], Fabio Gomez-Cano[a], Francisco M. Dillon[a], Roberto Alers-Velazquez[b], Andrea I. Doseff[c], Erich Grotewold[a], John Gray[d,*]

[a] Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI, 48824, USA
[b] Molecular Biology Graduate Program, University of Toledo, Toledo, OH, 43606, USA
[c] Department of Physiology, Department of Pharmacology and Toxicology, Michigan State University, East Lansing, MI, 48824, USA
[d] Department of Biological Sciences, University of Toledo, Toledo, OH, 43606, USA

ARTICLE INFO

Keywords:
Phenylpropanoid
Maize
Co-expression
Enzymatic module
Regulatory module
Regulomics

ABSTRACT

Phenolic compounds are among the most diverse and widespread of specialized plant compounds and underly many important agronomic traits. Our comprehensive analysis of the maize genome unraveled new aspects of the genes involved in phenylpropanoid, monolignol, and flavonoid production in this important crop. Remarkably, just 19 genes accounted for 70 % of the overall mRNA accumulation of these genes across 95 tissues, indicating that these are the main contributors to the flux of phenolic metabolites. Eighty genes with intermediate to low expression play minor and more specialized roles. Remaining genes are likely undergoing loss of function or are expressed in limited cell types. Phylogenetic and expression analyses revealed which members of gene families governing metabolic entry and branch points exhibit duplication, subfunctionalization, or loss of function. Co-expression analysis applied to genes in sequential biosynthetic steps revealed that certain isoforms are highly co-expressed and are candidates for metabolic complexes that ensure metabolite delivery to correct cellular compartments. Co-expression of biosynthesis genes with transcription factors discovered connections that provided candidate components for regulatory modules governing this pathway. Our study provides a comprehensive analysis of maize phenylpropanoid related genes, identifies major pathway contributors, and novel candidate enzymatic and regulatory modules of the metabolic network.

1. Introduction

Phenolic compounds, one of the most abundant families of specialized metabolites present in plants, have important functions in plant growth, development and adaptation. They are characterized by chemical structures that have at least one aromatic ring with one hydroxyl group, comprising more than 10,000 compounds so far identified in plants [1,2]. Among the phenolic compounds, phenylpropanoids are synthesized from the amino acids tyrosine and phenylalanine by a core and branch pathways that involved at least 25 enzyme families, and generate an enormous array of specialized metabolites that include the monolignols (which polymerize into lignins), the flavonoids (including anthocyanidins, proanthocyanidins, flavonols, flavonones, flavones, isoflavones and phlobaphenes), phenolic acids and stilbenes [3–5]. The pathways and enzyme families involved are summarized in Fig. 1 and Table 1.

Lignin is formed by the polymerization of monolignols [p-hydroxyphenyl (H), guaiacyl (G), syringyl (S) and tricin (T)] via oxidative radical-coupling reactions [6]. This heteropolymer is deposited in the secondary thickening of the cell walls [7]. Lignin, along with cellulose and hemicellulose, make up the bulk of the renewable biomass for the carbon-neutral production of biofuels and biochemicals [8]. Efforts to manipulate lignin accumulation have often been hindered by lignin modification-induced dwarfism, which might be associated with combinations of vasculature and cell wall integrity defects, and the accumulation of detrimental pathway intermediates [9,10]. Flavonoids are composed of two aromatic rings linked by a pyrone ring and they are classified into sub-classes based on the position and modifications of the A, B and C rings (Fig. 1). Flavonoids exhibit diverse functions in the physiology of plants, including attracting pollinators by the anthocyanins that pigment flowers [11], protecting against UV radiation and oxidative stress [12,13], serving as signals in plant development [14] and participating in plant-microbe [15] and plant-animal interactions [16].

* Corresponding author.
E-mail address: John.Gray5@utoledo.edu (J. Gray).

https://doi.org/10.1016/j.plantsci.2019.110364
Received 11 September 2019; Received in revised form 25 November 2019; Accepted 30 November 2019
Available online 05 December 2019
0168-9452/

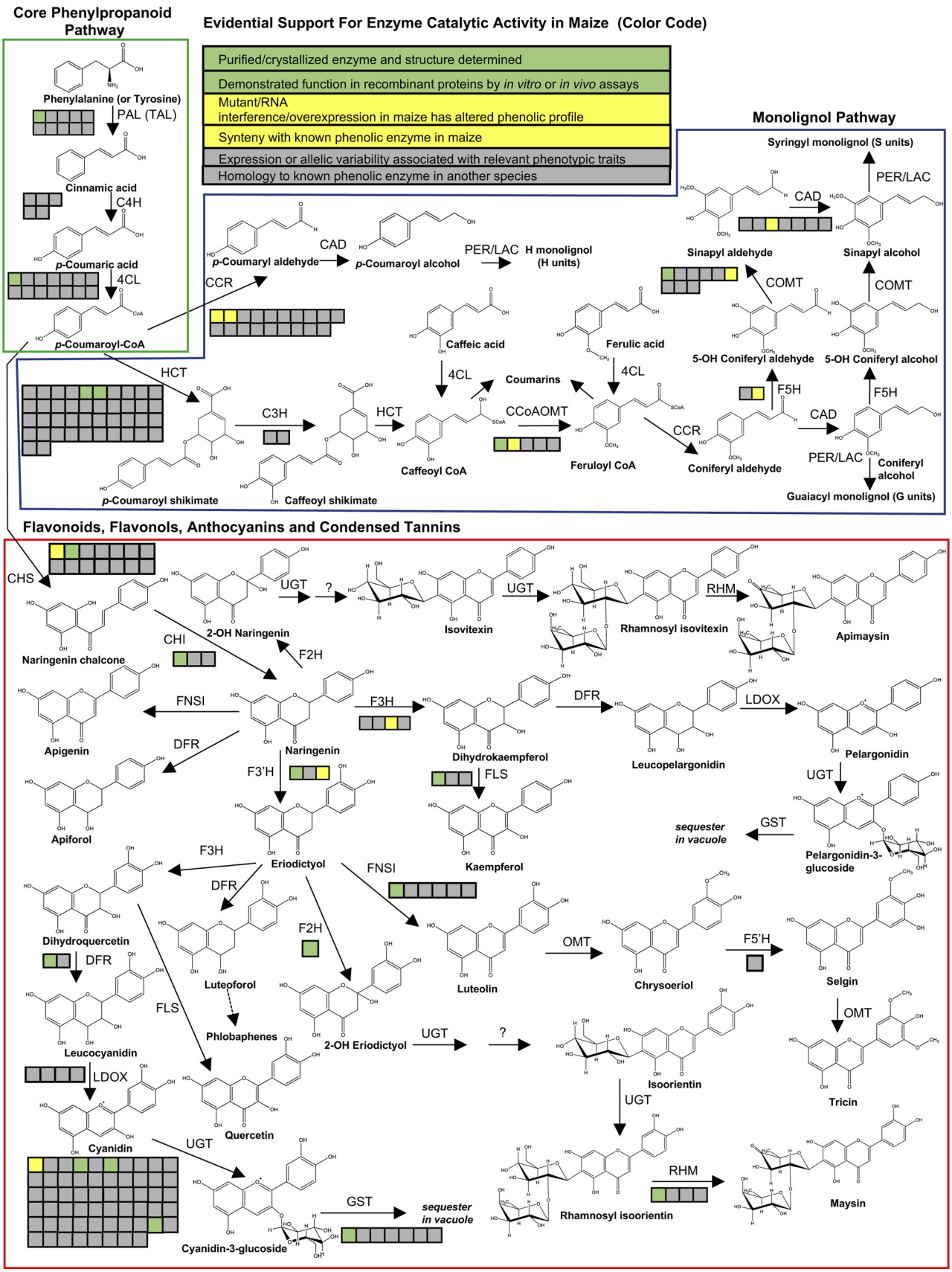

Fig. 1. Summary of Evidence for Core Phenylpropanoid, Monolignol, and Flavonoid Pathways in Maize. Boxes represent the estimated number of genes for a particular enzyme and are color coded to indicate the strength of evidence supporting the encoded enzyme activity (dark green: strong, yellow: medium and gray: weak). Enzyme abbreviations are provided in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of Phenylpropanoid and Flavonoid Biosynthetic Genes in Maize[a].

Enzyme	Full Name and Enzyme Type	EC number	No. in this study	Mutant in Maize	Reference
Core Phenylpropanoid pathway					
PAL	Phenylalanine ammonia lyase (aromatic amino acid lyase)	4.3.1.24	10		[112]
C4H	Cinnamate 4-hydroxylase (monoxygenase) (Cytochrome P450) (CYP73A)	1.14.14.91	5		[113,114]
4CL	4-Coumarate-CoA ligase (Acyl-CoA Synthetase)	6.2.1.12	14	*brown midrib 5*	[19,115,116]
CCR	Cinnamoyl CoA reductase (NAD(P)dependent epimerase/dehydratase)	1.2.1.44	19	*Cinnamoyl CoA Reductase 1*	[117,118,119]
		Total	48		
Monolignol pathway					
CAD	Cinnamoyl alcohol dehydrogenase (Zn-binding alcohol dehydrogenase)	1.1.1.195	7	*brown midrib 1*	[120,121,122]
HCT	Shikimate O-hydroxycinnamoyl transferase (Acyltransferase)	2.3.1.133	41		[87,97]
C3H	p-Coumaroyl shikimate 3-hydroxylase (monoxygenase) (Cytochrome P450) (CYP98A)	1.14.14.96	2		[114,123,124]
CCoAOMT	Caffeoyl CoA O-methyltransferase	2.1.1.104	5	*ccoaomt 1, ccoaomt 2,*	[125]
COMT/OMT	Caffeic acid O-methyltransferase	2.1.1.68/16	9	*brown midrib 3*	[126,127]
F5H (FAH)	Ferulic acid 5-hydroxylase (Cytochrome P450)	1.14.14.81	2		[20,114]
		Total	66		
Flavonoid pathway					
CHS	Chalcone synthase (Polyketide synthase, type III)	2.3.1.74	14	*colorless 2, white pollen 1*	[31,102,108,128]
CHI	Chalcone isomerase	5.5.1.6	3		[32,129]
F3′H	Flavonoid-3′-hydroxylase (flavonoid 3′-monooxygenase)	1.14.13.21	3	*purple aleurone 1*	[130,131]
FNSI	Flavone synthase I, (dioxygenase)	1.14.20.5	6		[34]
FNSII	Flavone synthase II (Cytochrome P450)	1.14.19.76	2		[37]
F3H	Flavanone 3-hydroxylase (dioxygenase)	1.14.11.9	4		[77,132]
F2H	Flavonoid-2-Hydroxylase (Cytochrome P450) (CYP93G5)	1.14.14.162	1		[133]
DFR	Dihydroflavonol 4-Reductase	1.1.1.219	2	*anthocyaninless 1*	[43]
FLS	Flavonol Synthase (dihydroflavonol,2-oxoglutarate:oxygen oxidoreductase)	1.14.20.6	3		[41]
ANR	Anthocyanidin reductase	1.3.1.77	3		[134]
LAR	Leucoanthocyanidin reductase	1.17.1.3	1		[135]
LDOX	Leucoanthocyanidin dioxygenase	1.14.20.4	4	[32,43,47,136]	
UGT	UDP glycosyltransferase	2.4.1.91	57	*salmon silks 2, bronze 1, ugt 14, ugt 46*	[54,137,138,139]
GST	Glutathione-S-transferase	2.5.1.18	7	*bronze 2*	[53,140]
AAT	Anthocyanin Acyltransferase	2.3.1.153	8	*anthocyanin acyltransferase 1*	[78]
RHM	UDP-Rhamnose synthase (NAD (P) dependent UDP-glucose 4,6-dehydratase)	4.2.1.76	4	*salmon silks 1*	[54]
		Total	122		
		Sum Total	236		

[a] The gene families *MTHFR*, *ALDH*, *PMAT* were not included in this table as they were considered peripheral to or predicted only to participate in the core, monolignol, and flavonoid pathways (Supp. Table 1).

The first committed step in the phenylpropanoid pathway is the conversion of L-phenylalanine to *trans*-cinnamic acid (CA) by phenylalanine ammonia lyase (PAL). A shortcut in the pathway is provided by the conversion of L-tyrosine into *p*-coumaric acid (pHCA) by tyrosine ammonia lyase (TAL). Monocot PALs also display TAL activity, and a recent study resulted in the identification of a *Brachypodium distachyon* bifunctional ammonia lyase responsible for the production of almost half of the lignin present in this model grass species [17]. The 4-hydroxylation of CA gives *p*HCA as product and is mediated by cinnamate-4-hydroxylase (C4H), which belongs to the cytochrome P450 family (CYP73A) [18]. The enzyme 4-coumarate:CoA ligase (4CL) exhibits distinct substrate specificities. It is involved in the third and last step of the general phenylpropanoid metabolism converting *p*-HCA into *p*-coumaroyl-CoA, and in the lignin pathway converting ferulic acid and caffeic acid to feruloyl-CoA and caffeoyl-CoA, respectively [19–21].

Hydroxycinnamoyl-CoA transferase (HCT) belongs to a large family of acyltransferases and esterifies *p*-coumaroyl-CoA or caffeoyl-CoA with shikimic or quinic acid [22]. HCT also is implicated in the reverse reaction by transforming caffeoyl shikimate to caffeoyl-CoA [22]. A similar reaction is catalyzed by caffeoyl shikimate esterase (CSE), which instead gives caffeic acid as product [23]. However, the corresponding gene has not been definitely identified yet in maize. Cinnamate-3-hydroxylase (C3H) is a cytochrome P450 (CYP98A3) that catalyzes the 3-hydroxylation of shikimic or quinic acid-esterified *p*-HCA into their corresponding caffeic acid conjugates [24].

Caffeoyl-CoA O-methyltransferase (CCoAOMT) methylates the 3-OH of caffeoyl-CoA to make feruloyl-CoA [25]. Cinnamoyl CoA reductase (CCR) catalyzes the conversion of hydroxycinnamoyl-CoA esters (*p*-coumaroyl-CoA and feruloyl-CoA) to their corresponding cinnamyl aldehydes (coumaryl and coniferyl aldehydes) [26]. Subsequently, cinnamyl alcohol dehydrogenase (CAD) catalyzes the reduction of *p*-hydroxycinnamaldehydes into their corresponding alcohols [27]. Either coniferyl aldehyde or coniferyl alcohol are substrates of coniferaldehyde 5-hydroxylase (F5H;) [28], which is a cytochrome P450-dependent monooxygenase (CYP84) that converts them to the respective 5-OH derivatives. The 5-hydroxylation can be followed by a 5-methylation catalyzed by a caffeic acid *O*-methyltransferase (COMT), which also uses aldehydes or alcohols as substrates to make sinapyl aldehyde or sinapyl alcohol [29]. The monolignols *p*-coumaryl, conferyl and syringyl alcohols are polymerized via oxidative radical coupling reactions to form, respectively, the H, G, and S units of lignin [30].

The activated phenylpropanoid *p*-coumaroyl CoA can be used either for lignin biosynthesis (as described above), or for the formation of flavonoids by chalcone synthase (CHS), through conjugation with malonyl-CoA to form naringenin chalcone. In maize, *C2* (*colorless2*) and *WHP1* (*white pollen1*) are duplicate loci encoding the enzymes responsible for this entry step into the pathway [31]. Chalcone stereospecific isomerization is facilitated by chalcone isomerase (CHI), which

leads to the formation of the flavanone naringenin [32]. Flavonoid-3′-hydroxylases (F3′H) belong to the cytochrome P450 protein family and hydroxylate naringenin in ring B to form eriodictyol [33]. The same enzyme catalyzes the transformation of (dihydro)kaempferol to (dihydro)quercetin [33].

The flavone synthase I (dioxygenase) enzyme (FNSI) generates a double bond in ring C of the flavonones naringenin and eriodictyol to produce the flavones apigenin and luteolin, respectively [34,35]. The same oxidation can be done by the flavone synthase II (cytochrome P450 hydroxylases) enzyme (FNSII), which converts the flavonones into flavones [36,37]. Flavanone-3-hydroxylase (F3H) hydroxylates the flavanones naringenin and eriodictyol at the 3 position of the C-ring, resulting in the formation of the dihydroflavonols, dihydrokaempferol (DHK) and dihydroquercetin (DHQ), respectively [38]. F3H is important for the production of anthocyanins and flavonols [39], the latter being essential for pollen fertility in maize and other plants [40]. Flavonols are produced by a flavonol synthase (FLS), which is a dioxygenase enzyme that directs the oxidation of the DHK and DHQ to form kaempferol and quercetin, respectively [41,42].

The dihydroflavonol 4-reductase (DFR), encoded by *A1* in maize [43], catalyzes the reduction of the carbonyl group of the ring C of eriodictyol and dihydroquercetin to form respectively 3-deoxy flavan 4-ols (luteoforol, which then polymerize into phlobaphenes), and 3-hydroxy flavan 4-ols (leucocyanidin; which can undergo subsequent reactions to lead to cyanidins) [44]. The anthocyanidin synthase (ANS) enzyme, also termed leucoanthocyanidin dioxygenase (LDOX), is involved in the production of cyanidin from leucocyanidin [45,46], and is encoded in maize by *A2* [47,48]. In some plants, cyanidin can be a substrate of anthocyanidin reductase (ANR) to produce flavan 3-ols [49].

Flavonoid stability, solubilization and subcellular compartmentalization depend on the conjugation with sugars and acyl groups [50]. UDP-glycosyltransferases (UGTs) form a large group of enzymes that transfer sugar moieties from activated sugar donors to a wide range of molecules, including flavonoids and anthocyanidins [51]. Malonylation is the most common acylation of flavonoid glucosides and is mediated by malonyl transferases [52]. The subcellular transport of anthocyanidin glucosides is believed to depend on Type III glutathione *S*-transferases (GSTs) that reversibly bind to them [53]. Although it is less commonly reported, sugar moieties from flavonoid glycosides can undergo irreversible modifications. Such is the case of rhamnosylisoorientin, whose glucose directly attached to the flavone is converted to deoxy-keto glucose by a rhamnose synthase, to form maysin [54].

While the phenylpropanoid pathway has been extensively investigated in dicots such as poplar and *Arabidopsis*, significant gaps remain in the grasses. This is particular the case in maize, where most of the pathway genes are part of small gene families with likely partially redundant functions, making it particularly challenging to determine which specific family member is predominantly involved in pathway flux. Even though many flavonoid genes were first characterized in maize thanks to the conspicuous effects of the mutants on pigmentation [55], no mutants have yet been found for chalcone isomerase (CHI) or flavanone 3-hydroxylase (F3H).

The regulation of the pathway is complex and involves conserved sets of transcription factors that control specific pathway modules [56]. For example, R2R3-MYB transcription factors were among the first to be identified as directly activating pathway and other transcription factor genes [57]. Several of these R2R3-MYBs are downstream targets of members of the VNS protein (for VND, NST/SND, and SMB-related protein) subfamily, which have been extensively characterized in dicots [58]. A distinct class of R2R3-MYBs are gaining importance as pathway repressors [59]. Efforts to identify similar regulatory components in maize have involved the generation of an extensive transcription factor open reading frame (ORF) set [60], that was subsequently used in yeast one-hybrid (Y1H) experiments to identify over 1000 novel protein-DNA interactions involving the regulatory regions of 54 maize phenylpropanoid genes and 500+ transcription factors [61]. Analysis of high-throughput RNA-seq data can give empirical information about how gene expression of enzymes from the same biosynthetic pathway is associated or coordinated to generate a carbon flux in the proper direction.

Here, we describe a comprehensive analysis of the maize phenylpropanoid related genes. Previous attempts to catalog all genes linked to phenylpropanoid metabolism in *Arabidopsis* revealed 65–80 candidate genes of which just 13 could be classified as having a definitive function [62,63]. By combining comparative genomics with expression analysis, 15 *bona fide* genes involved in lignin biosynthesis could be identified in poplar [64]. A genomic survey of the genes linked to cell wall biosynthesis in maize included 102 linked to phenylpropanoid metabolism [65]. Here, we used available genome-wide expression analyses to determine which family members are predominantly expressed, and explored their co-expression with other pathway genes as a first step towards establishing which isoform assemblies might participate in the formation of the various pathway products. We also explored the patterns of gene duplication, and determined, in some families, paralog function using co-expression analyses. Finally, we analyzed available protein-DNA interaction datasets and, combined with co-expression analyses, predicted which transcription factors might be involved in the control of the various enzyme gene modules. Together, these studies provide a valuable framework to further advance studies on the biosynthesis and control of phenylpropanoids in maize and other grasses.

2. Materials and methods

2.1. Identification of candidate maize phenylpropanoid genes

The selection of genes linked to maize phenylpropanoid metabolism consisted first, of a literature search to identify 47 genes from Arabidopsis already validated to play a role in phenolic metabolism (Supp. Table 2). Next, candidate maize gene orthologs were identified by performing a BLASTP search against the maize B73 v4 genome available from the Gramene database (http://www.gramene.org/) with a homology cutoff ≥ 90 %. The MaizeGDB database (http://www.maizegdb.org) was used to help identify maize genes that have been well-characterized as being linked to phenolic metabolism (Table 1 and Supp. Table 1). Synteny between maize and sorghum gene models was as determined previously using SynMap [66] and intragenomic maize syntelogs were found using the SynFind algorithm (Supp. Table 1) [67].

2.2. Generation of phylogenetic trees

Protein sequences were aligned using the MUSCLE algorithm as implemented within the MegAlign Pro program (DNASTAR®, Madison, WI). Phylogenetic trees that were created based on multiple sequence alignments (MSA) using default parameters were exported to Figtree (V1.4.4) for graphical display [68]. The examination of MSAs generated by the MUSCLE algorithm revealed discrepancies in several gene models. To resolve gene model uncertainties, the predicted transcripts were checked against existing RNA-seq datasets and corrected based on read support for intron-exon junctions. The RNA-seq datasets employed for this purpose were from nine expression studies (66 maize tissue samples) publicly available from the qTELLER database (www.qteller.com). The raw datasets were downloaded in the FASTQ format and mapped to the maize B73 reference genome (v4 rel. 37) using the HiSat2 Linux package (v2.1.0) [69]. The Integrated genome viewer (IGV) was utilized to visualize the gene expression data and when mapped reads indicated an error in the gene model, a corrected transcript was used for conceptual translation and MSA (Supp. Table 1).

2.3. Gene expression and co-expression analyses

The expression levels of core phenylpropanoid, monolignol, and flavonoid genes in the maize B73 inbred reference germplasm were analyzed using previously published RNA-seq data from the B73 genome and included samples from 95 tissue samples [70]. The RPKM values for the selected genes were retrieved, and the values for developmental or stress induced samples were normalized by transforming to $\log_2$ RPKMs. A heat-map of the tissue-specific expression pattern was established using the R package Hmisc [71]. Co-expression was calculated as the mutual rank (MR) of the mutual information index (MR-MI) between all gene models. Significant MR-MI values were defined as the top 5 % of the MRs. The MI was estimated from 3-nearest neighbors [72] using the R package Parmigene (v1.0.2) [73], and with $1e^{-12}$ as noise to break ties due to limited numerical precision. An alternative approach to detect co-expression based functional regulatory modules used the same data source and employed average hierarchical clustering analysis, with Euclidean distances as implemented within the R package pvclust [74]. The uncertainty in hierarchical cluster analysis was assessed by multiscale bootstrap resampling (n = 10,000) which provided approximately unbiased (AU) *p*-values, as well as bootstrap probability (BP) values for each cluster in a dendrogram. Clusters with AU values < 95 were considered uncertain.

3. Results and discussion

3.1. Assigning genes to the maize phenylpropanoid enzymes

We utilized the repertoire of *Arabidopsis* phenylpropanoid and flavonoid metabolism genes to identify candidate orthologs (homology ≥ 90 %) in the maize B73 (v4) genome. The protein sequences of these candidate orthologs were then used to search the maize genome for similar proteins using BLASTP and a total of 249 gene models from 28 gene families were identified (Supp. Table 1). The methylenetetrahydrofolae reductase (MTHFR) and aldehyde dehydrogenase (ALDH) families were considered to have indirect roles in the pathway and were omitted from further analysis. The remaining 26 families were developed for further expression and coexpression analyses (Table 1 and Supp. Table 1). Some families such as *UGT*, *HCT*, *CCR*, *4CL*, and *CHS* had a relatively larger number of gene models included (58, 42, 19, 14, and 14 models, respectively) because they exhibit low homology outside a conserved core motif. These were included in this study out of a preference to be inclusive until substrate specificity is investigated. The overall criteria for inclusion of each gene model resulted in a classification of 'weak', 'medium' or 'strong', depending on various criteria, including homology, expression, and biochemical activity obtained in this study or from the existing literature (Table 1, Fig. 1, and Supp. Table 1). Gene models that exhibited homology to a known phenolic enzyme in another species, or whose expression was associated with relevant phenotypic traits, were classified as weak. Syntenic genes with a known phenolic enzyme gene in maize or for which a mutant, RNA interference, or overexpression results in maize exhibited an altered phenolic profile were classified as medium. Gene models with a demonstrated enzymatic function by *in vitro* or *in vivo* assays, or for which structure had been determined, were classified as strong. We found strong evidence that members of 17 enzyme families participate in the phenylpropanoid, monolignol, and flavonoid pathways in maize, whereas there was medium evidence for four more families (Table 1, Fig. 1, and Supp. Table 1). Of these 17 families, there are known mutations in members of at least 12 of them, many of which were first described from classical genetic studies in maize and later characterized at the molecular and biochemical level (*e.g., c2* and *whp,* and the *brown midrib* mutants *bm5*, *bm1*, and *bm3* that encode 4CL, CAD and COMT, respectively) (Table 1).

3.2. A subset of biosynthesis genes plays a predominant role in phenolic metabolism

3.2.1. Identification of major contributing genes in phenolic metabolism

Our survey of the maize genome uncovered a repertoire of 249 candidate gene models possibly participating in phenolic metabolism. We hypothesized that only a subset of these genes would play a predominant role, and that these could be identified by determining those that exhibit a high level of expression across a large number of maize tissues. For this purpose, we profiled the expression level of 205 candidate genes across 95 tissues samples previously characterized by RNA-seq [70]. We also omitted eight *PAL* pseudogenes and some members of large families such as HCT (HCT Clades 1, 2 and 3) that were more distantly related from family members whose phenylpropanoid function was strong (*i.e.* HCT Clade 4), hence we analyzed 205 of the 249 candidates. Assumptions in these analyses are that mRNA levels can reflect protein levels and that mRNA expression levels can be correlated with the accumulation of end products of a biosynthetic pathway [75]. The overall average expression level of all genes examined ranged from 1172 FPKM (ZmGST1) to 0 (several) with an average of 49 ± 249 FPKM (Supp. Table 3). Remarkably, we found that 8–19 phenolic genes contribute to 50 %–70 % of all the total reads (FPKM) for the 205 genes across 95 tissues with an average FPKM > 122 (Fig. 2). This unexpected observation suggests that a small gene subset are predominantly expressed to facilitate the flux of intermediates through the various branches of the phenolic pathway. *ZmPAL1* and *ZmC4H1* are in the top eight, suggesting that these enzymes funnel most intermediates into the pathway. Likewise *ZmCCoAOMT1*, *ZmCOMT1*, and *ZmOMT8* appear as major contributors to the flux through the monolignol pathway. *ZmGST1*, *ZmGST4*, and *ZmUGT45* are also in the top predominantly expressed genes, indicating that enzymes facilitating the final steps of glycosylation and stabilization/glutathione conjugation that contribute to transport into the vacuole also may play a predominant role. Eighty two genes exhibited an intermediate level of expression (average FPKM between 10 and 121) and are likely to represent genes whose products are either required in lower abundance, in a few cell types, or that play more specialized roles in phenolic metabolism. At the low end of the expression spectrum there were 75 genes with an average FPKM between 1 and 10, and

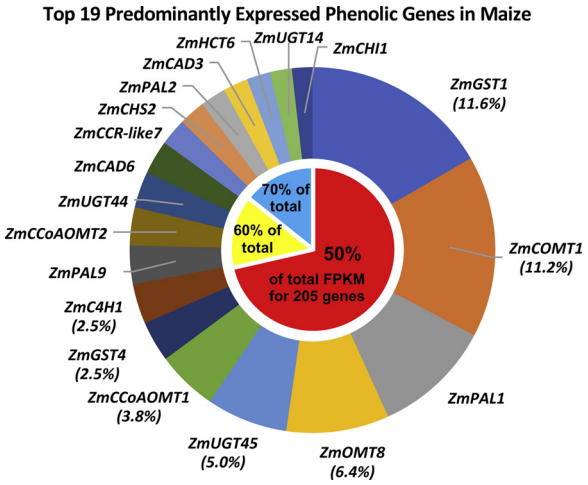

Fig. 2. Summary of Most Predominantly Expressed Maize Phenolic Genes. Outer pie chart summarizes the relative overall expression level across 95 tissues of the indicated phenolic genes in maize (dataset from Ref. [70 of the top 19 predominantly expressed genes. These 19 genes collectively contribute 70 % of the expression (% of total FPKM) of the 205 genes examined (956,805 total FPKM). The individual overall contribution of the eight genes that collectively provide 50 % of the overall read count are shown in brackets. A full list of average expression levels of all 205 genes is provided in Supp. Table 3.

further 29 with an average FPKM < 1. The latter genes which collectively contribute < 4 % of the total FPKM value for all 205 genes) might be expected to contribute little to the overall functioning of the pathway, although some may have a specialized role in certain developmental tissues such as *Zm4CL-like9* which is expressed almost uniquely in meiotic tassel tissue (FPKM 358).

In summary, our comprehensive analyses listed 205 candidate genes from 26 gene families and revealed that just 19 genes accounted for 70 % of the overall mRNA accumulation across 95 tissues. Our interpretation is these 19 genes produce the main enzyme complexes that function at key entry, branch, and exit points, to control metabolic flux. For 12 of these gene families, classical mutants are already available. The fact that only some of these genes were identified by classical mutant studies may be due their essential function in maintaining plant growth and/or the availability of a duplicate gene to compensate when a mutation occurs. The availability of new gene editing technologies should permit non-lethal allelic series to be developed for the investigation of genes for which mutants have not yet been discovered (e.g. *PAL* and *CHI* family members). From an application viewpoint, this set of ~40 phenylpropanoid genes represent those that are most likely to be targeted for trait manipulation. It will be of interest to determine the extent to which the predominant role of certain genes is conserved also in other inbred lines.

3.2.2. Developmental expression profiles of predominantly expressed core phenylpropanoid and monolignol genes

To further investigate the roles that different phenolic genes might play, we examined the developmental expression profile of the different gene families in the core phenylpropanoid and monolignol pathways (Fig. 3) and in the flavonoid pathway (Fig. 4). A published dataset of 95 tissue samples included 79 tissue samples from various stages of development and 16 tissue samples under stress conditions [70]. Overall, we found that the core phenylpropanoid genes *PAL*, *C4H* and *4CL* are expressed in leaves peaking around the V9 stage of development, before declining in older, fully developed leaves. These core genes were expressed highly in most stem and root tissues, which could be expected as these tissues are more heavily lignified (Fig. 3). In flowers, the core enzymatic genes are expressed in tassels and silks, but at low levels in the immature/unfertilized cob. Within developing seeds, expression levels of core enzyme genes was highest in the first two weeks after pollination and declined thereafter. Expression in the endosperm was very low, but expression in the pericarp (18 DAP) was notably high as is expected due to the protective nature of this outer layer of maternal tissue (Fig. 3). As hypothesized, a subset of each gene family was also seen to be expressed in a tissue specific fashion. For example, *ZmPAL1,9,2,5* and *3* are expressed at much higher average levels (251 ± 283 FPKM) than *ZmPAL4,6,7,8* and *10* (14 ± 12 FPKM). Likewise, *ZmC4H1* and *3*, *Zm4CL1,2* and *5* and *Zm4CL-like1,2*, and *4* play predominant roles in their respective families (Fig. 3). In conclusion, our results suggest that the majority of the metabolic flux through the core phenylpropanoid pathway in maize is conducted via *ZmPAL1,9*, and *2, ZmC4H1*, and *Zm4CL1*.

Within families of the monolignol pathway, subsets of genes were also predominantly expressed (*ZmCCR1* and *2*, *ZmCCR-like7,8* and *3*, *ZmHCT6,15,9* and *11*, *ZmCAD6,3,7* and *2*, *ZmCCoAOMT1* and *2*, *ZmCOMT1* and *ZmOMT8*). In the *C3H* and *FAH* families, the overall contribution of each of the two identified members (which are syntelogs) differed in a tissue-specific fashion, likely reflecting subfunctionalization of syntelogs. For example, *ZmC3H1* is expressed at much higher levels in leaves than *ZmC3H2,* and *ZmFAH1* is expressed at higher levels than *ZmFAH2* in stems (Fig. 3). The predominant members of the monolignol metabolism genes were also seen to be expressed in lignifying V9 leaves, internodes, roots and in drought stressed roots (Fig. 3). In summary, based on predominant expression levels, it appears that the majority of the metabolic flux through the monolignol pathways is conducted by a specific gene subset controlled in a tissue-specific fashion.

3.2.3. Developmental expression profiles of predominantly expressed flavonoid genes

We also examined the developmental expression profiles of flavonoid metabolism (Fig. 4). The B73 inbred line harbors a *P1-wr, pl1, and b1* alleles and accumulates lower levels of phlobaphenes than many other lines [70,76]. Flavonoid genes were expressed throughout most of plant development, but at highest levels in older leaves, roots and under drought and biotic stress (Fig. 4). *ZmCHS2* (*C2*) clearly plays the predominant role in the entry reaction of naringenin chalcone biosynthesis across most tissues with a lesser role for the duplicate gene *ZmCHS1*(*Whp1*) in leaves, tassels and anthers, and notably during biotic infection (Fig. 4). Interestingly, *ZmCHS14*, *13, and 11,* are expressed almost exclusively in meiotic tassel tissue, suggesting their specialized role in pollen. *ZmCHI1* (average 128 ± 187 FPKM) was the predominantly expressed gene in the second step of the flavonoid pathway with *ZmCHI2* (average 75 ± 135 FPKM) and *ZmCHI3* (average 29 ± 52 FPKM) playing lesser, but largely overlapping roles, possibly explaining the lack of a mutant in the classical gene set. *ZmFNS1* plays the predominant role in the branch that synthesizes apigenin from naringenin and is expressed in older leaves and in leaves under biotic stress (Fig. 4). In the conversion of naringenin to eriodictyol, *ZmF3′H1* (*Pr1*) plays the predominant role with a lesser role for *ZmF3′H2*, although the latter is expressed more in drought stressed roots and the former in leaves under biotic stress. *F3H* enzymes can participate in two branches of the flavonoid pathway in the conversion of eriodictyol to dihydroquercetin and in the conversion of naringenin to dihydrokaempferol (Fig. 1). The *F3H-like* genes were expressed at relatively low levels, but *ZmF3H1* [77] was expressed at relatively high levels in brace roots, anthers and in drought and biotic stressed tissues (Fig. 4). Amongst the flavonoid reductases (*DFR*, *ANR* and *LAR*), the classical gene *ZmDFR1*(*A1*) was expressed at low levels throughout development (average FPKM of 6.05) likely reflecting the recessive nature of the anthocyanin regulators in this inbred, but importantly it was expressed in root tissues where the duplicate *ZmDFR2* (*A4*) was not (Fig. 4). *ZmDFR2* (*A4*) was expressed at low levels in leaves, at intermediate levels in developing seeds, and notably high levels in brace roots (877 FPKM) and in silks (4308 FPKM). *ZmLDOX1* (*A2*) contributed the most to the cyanidin biosynthetic branch of the pathway and was expressed at notably higher levels in brace roots (230 FPKM) and in drought stressed roots (428 FPKM). The largest family of genes that was studied was the UDP glycosyltransferases (UGTs) with 58 members. *ZmUGT44* and *45* (which are syntelogs) and *ZmUGT14* were amongst the top eight predominantly expressed genes (Figs. 2 and 4) being highly expressed throughout the plant. They also exhibited high expression in meiotic tassel tissue (pollen production) and in silks. Eleven UGTs (*ZmUGT20, 26, 3, 41, 47, 29,13,18,16, 54* and *51*) exhibited expression more specifically in older leaves and abiotic and biotic-stressed leaves. Also, it was noted that *ZmUGT11* and *4* (*Sm2*) were expressed quite specifically in young leaf tissue and in silks. *ZmUGT4* (*Sm2*) encodes a rhamnosyl transferase that, in concurrence with *Sm1* (RHM, UDP-rhamnose synthase), are responsible for the production of maysin which protects the silk from herbivorous insects [54,75]. Within the *AAT* family, it was found that *ZmAAT1*, which has been linked to anthocyanion acylation [78] was predominantly expressed in roots under drought and biotic stress (Fig. 4). The homolog *ZmAAT2* was expressed at similar levels but also in leaf and root tissue. Lastly, it was found that amongst the GSTs examined, that *ZmGST1* and *2* were highly expressed throughout the plant in nearly all tissues and were amongst the top eight predominantly expressed phenolic genes. Some GSTs have been reported as required for transport of flavonoids into the vacuole and their high expression may reflect their important role in ensuring correct compartmentalization of phenolic compounds [50]. *ZmGST2* played a more specialized role with expression largely restricted to root tissues and drought stressed roots (Fig. 4).

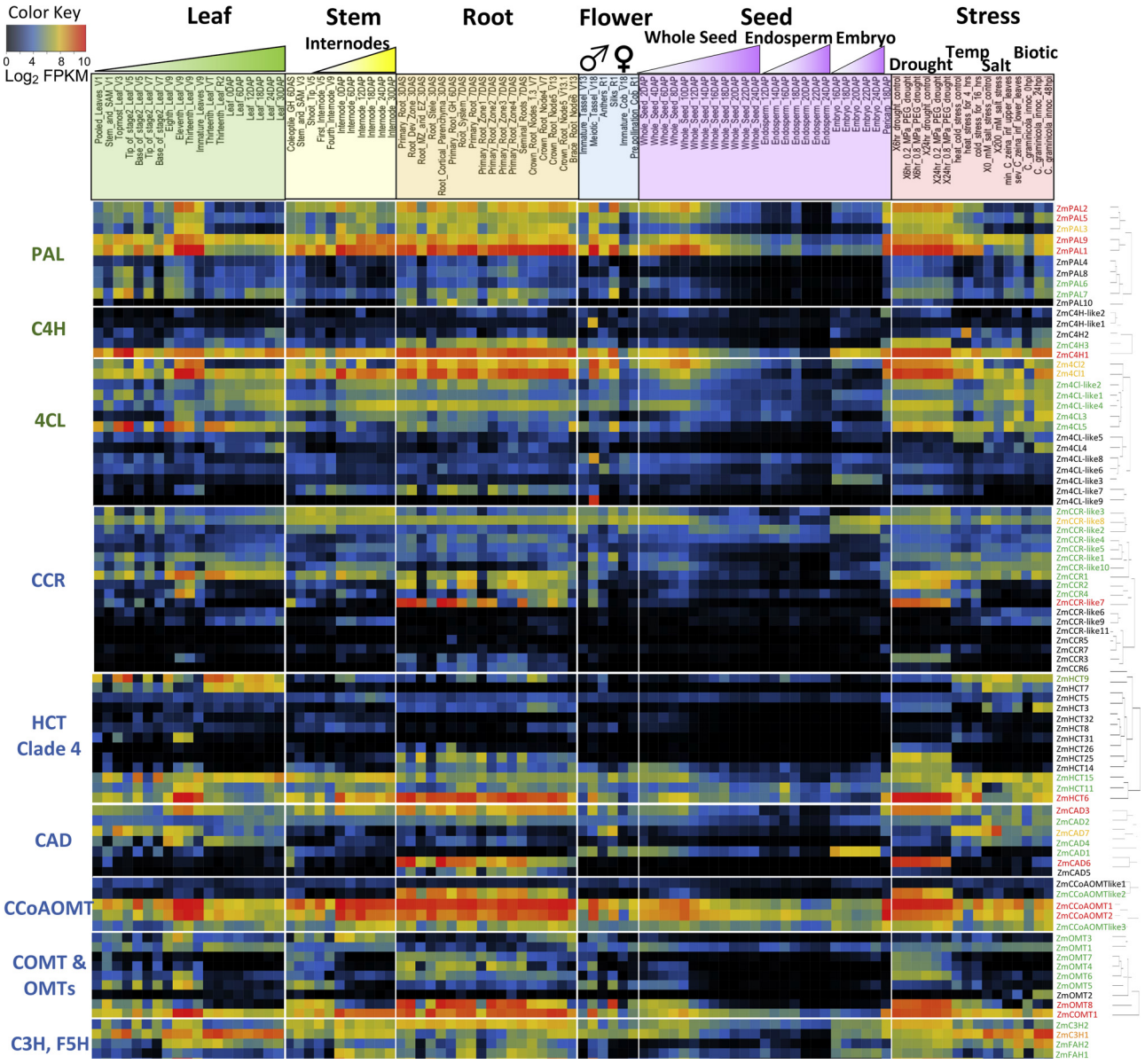

Fig. 3. ***In silico*** **Expression Analysis of Phenylpropanoid and Monolignol Genes in Maize.** Heat map representation of tissue-specific expression levels (Log_2 FPKM) of phenylpropanoid gene organized by enzyme family (dataset from Ref. [70]). Gene names within each family that exhibit a predominant role in phenylpropanoid metabolism are colored red if expressed at high levels (average FPKM > 122, *i.e.* top 19 most highly expressed), orange (FPKM 50–121), green (FPKM 10–49) and black (FPKM < 10). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In summary, our analyses established that a subset of flavonoid metabolism genes seem to be responsible for the majority of the flux through the various branch pathways, and that some family members exhibit more restricted developmental roles. The fact that *ZmCHS2*(*C2*), *ZmCHI1*, *ZmFNSi1*, *ZmFHT1*, *ZmDFR2* (*A4*), *ZmLDOX1* (*A2*), *ZmUGT44,45*, and *ZmGST1,4* exhibit some of their highest levels of expression in stressed tissues (Fig. 4) underscores the fact that the entire pathway is also responding to stress signals, but notably we did not observe family members that were uniquely stress-induced.

3.3. Phylogenomic analysis of maize phenylpropanoid genes

Our earlier results revealed that, even in large enzyme families, only a small subset of members, appear to play a major role in metabolizing the relevant substrates, suggesting that other members have either a reduced function, specialized function or lost function. Since the maize genome has undergone a recent whole genome duplication (WGD) event since its divergence from sorghum, we investigated the evolution of some of these gene families in an effort to discern patterns of gene subfunctionalization or loss (pseudogenization).

3.3.1. Maize PAL gene repertoire - entry point to the phenylpropanoid pathway

The *PAL* gene family in maize was found to contain 10 members. We took advantage of the sorghum genome to identify syntelogs and homologs of the maize *PAL* genes and used the aligned proteins to generate a phylogenetic tree (Fig. 5A). One gene model (Zm00001d033286, *ZmPAL10* on chr 1) had not been included in previous studies of PALs in maize [65] but appears to be functionally intact encoding the cofactor 3,5-dihydro-5-methylidene-4H-imidazol-4-one (MIO) binding, shielding, and core domains [79]. In maize there are three PAL clusters on chr 2, 4 and 5 (Fig. 5C). Overall, the phylogenetic relationship amongst the 10 intact maize *PAL* genes supports the interpretation that most duplication and clustering events occurred

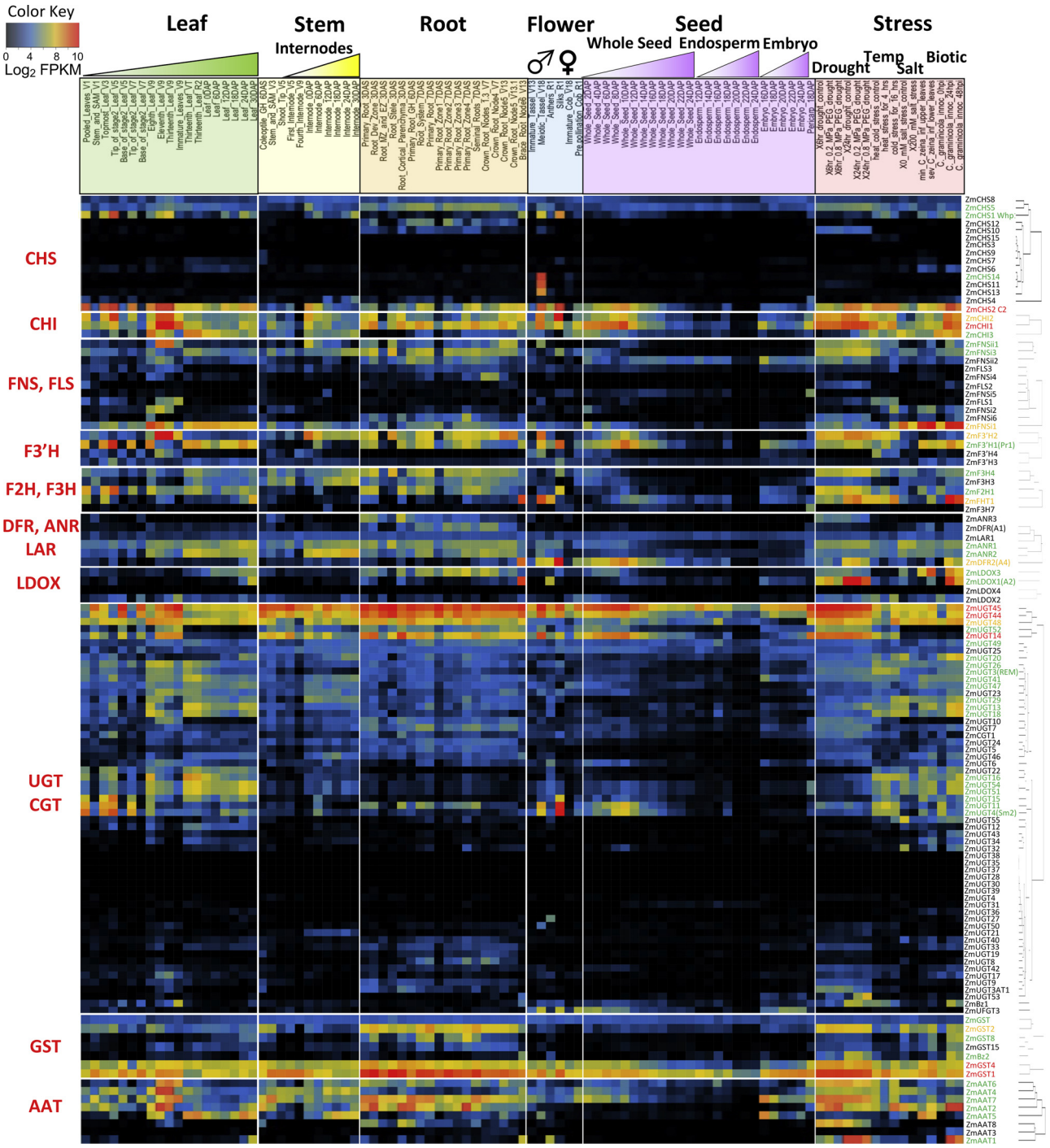

Fig. 4. *In silico* Expression Analysis of Flavonoid Metabolism Related Genes in Maize. Heat map representation of tissue-specific expression levels (Log_2 FPKM) of flavonoid and flavonol genes (dataset from Ref. [70]). Gene names within each family that exhibit a predominant role in flavonoid metabolism are colored red if expressed at high levels (average FPKM > 122 *i.e.* top 19 most highly expressed), orange (FPKM 50–121), green (FPKM 10–49) and black (FPKM < 10). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

prior to the divergence of sorghum and maize [80]. A cluster of 5 *PAL* genes on sorghum chr 4 is similar to the organization of the cluster of four PAL genes (*ZmPAL1,9,8* and *7*) on maize chr 5 (Fig. 5C), except for the absence of one of the single exon *PAL* genes. The two *PAL* genes located together on sorghum chr 6 are also organized similarly to the two *PAL* genes (*ZmPAL6* and *2*) on maize Chr 2 (Fig. 5C), and a single *PAL* gene on sorghum Chr 1 is most similar to *ZmPAL9* on maize chr 1. The cluster of *PAL* genes on maize chr 4 (*ZmPAL3, 5, 4* and a pseudogene) (Fig. 5C), is organized similarly to the cluster on maize chr 5 and appears to be a result of the recent WGD event in maize. These clusters are not the result of recent intracluster tandem repetition, however, since *ZmPAL1*, *3*, and *2*, on 3 different chromosomes are most homologous to each other (Fig. 5A, C). Likewise *ZmPAL9, 5, 6*, and *ZmPAL4, 8, 7*, cluster together while *ZmPAL10* appears distinct from the other PALs (Fig. 5A,C). The multiple sequence alignment of maize PAL proteins reveals that these three sets of homologs encode His, Phe, and Tyr respectively at residue 123 (of *ZmPAL1*). A biochemical study of *SbPAL1* indicates that His123 permits bifunctional PAL and TAL activity, whereas the presence of Phe or Tyr permits only PAL or very low PAL activity respectively [79]. Therefore, these clusters contain

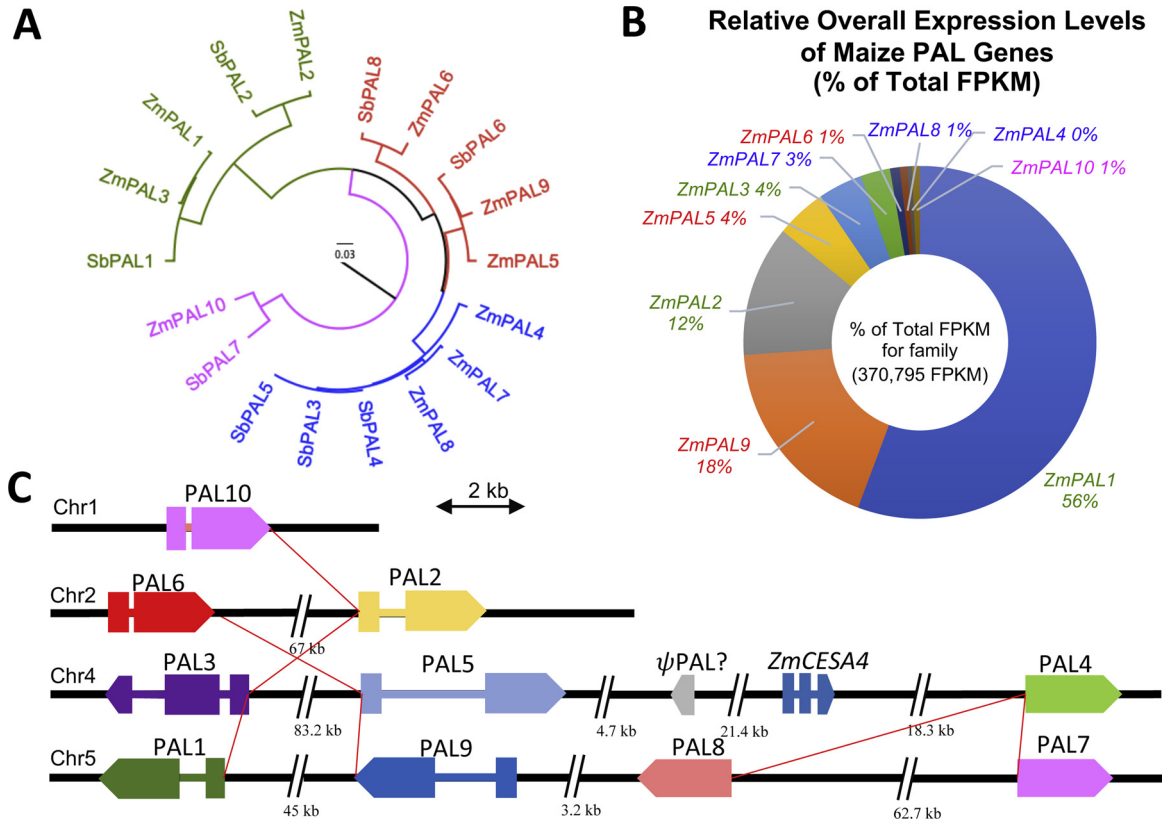

Fig. 5. Phylogenetic Relationships, Expression Analysis, and Gene Organization of Maize *PAL* Genes. A: Phylogenetic relationship amongst maize and sorghum PAL genes. **B:** Pie chart of relative overall expression levels of maize B; PAL genes, expressed as a % of total FPKM for the family across 95 tissues (dataset from Ref. [70]) **C:** Scaled representation of maize PAL gene clustering in the maize genome. Red connecting lines indicate phylogenetic proximity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

intragenomic syntelogs with potentially differing substrate preferences and thus, it can be hypothesized that these PAL genes might exhibit a pattern of coordinate expression within or between different clusters in the genome (see below). Although there are 10 family members in maize, it is clear that *ZmPAL1* (which is possibly bifunctional) and *ZmPAL9* are predominantly expressed and account for 56 % and 18 %, respectively, of the mRNA accumulation attributable to this family across 95 maize tissues (Fig. 5B). Within the same cluster on chr5, *ZmPAL8* is almost silent suggesting pseudogenization of this family member. *ZmPAL2*, which is possibly bifunctional, (12 %), *ZmPAL5* (4.6 %) and *ZmPAL3* (4 %) account for a smaller contribution but their expression pattern largely parallels that *of ZmPAL1* and *9* (Fig. 3). *ZmPAL7*, at the end of the cluster on chr 5, contributes 2.7 % of PAL activity (deduced from mRNA accumulation levels), mainly in leaves and roots. The remaining four PAL members *ZmPAL6,8,4* and *10* collectively contribute only 3 % of overall reads for this enzyme family and the latter appears restricted mostly to expression in root tissues. Collectively these observations point to major roles for both *ZmPAL1* and *9* on chr5 and diminishing roles for their orthologs *ZmPAL2* and *3* and *ZmPAL5* and *6* respectively in the core phenylpropanoid pathway.

3.3.2. Maize 4CL gene repertoire - exit point from the core phenylpropanoid pathway

To investigate genes regulating the exit point from the core phenylpropanoid pathway, predicted *4CL* and *4CL-like* encoding genes from maize and sorghum were aligned and used to create a phylogenetic tree (Fig. 6A). We identified five *4CL* genes in both species, a similar number reported for *Arabidopsis* and rice [19,23,81,82]. The conservation of the number of *4CL* clade members occurs despite the fact that a WGD event in maize has occurred following divergence from sorghum. Of these, *Zm4CL1* and *2* on chr 5 and 9 are syntelogs and are expressed at the highest levels in maize tissues (Fig. 6B). The five proteins encoded by the 4CL clade exhibit conserved amino acids in the substrate binding pocket as found in *Sb4CL1* (*Bmr2*) and related 4CL enzymes [81,83,84]. The proteins of the 4CL-like clades conserve most, but not all of these residues, suggesting the possibility of similar substrate specificity. In support of this hypothesis, we found that the sister *4CL-like* clade-1 exhibited seven members of which *Zm4CL-like3* is most similar to the *Arabidopsis* gene *At4g19010* (Fig. 6A). *At4g19010* encodes a 4CL implicated in diverting *p*-coumarate from the phenylpropanoid pathway to ubiquinone biosynthesis [85,86]. The existence of the 4CL-like clade-1 in maize suggests that this ubiquinone biosynthesis pathway may also function in monocots. In support of this hypothesis, it was found that all seven members of the 4CL-like1 clade exhibit stringent conservation of a C-terminal S-R/K-L consensus motif (data not shown) known to target proteins to the peroxisome where *At4g19010* has been shown to localize [85]. Expression analysis shows that, within the *4CL-like* Clade-1, that *Zm4CL-like 4*, *2* and *5* are predominantly expressed, yet at low levels (Fig. 6B). In summary, phylogenetic analyses show the conservation of just five *4CL* members in maize, resembling results found in other monocots and dicots [19,81,82] and the existence of several *4CL-like* genes that may contribute to the channeling of *p*-coumarate to ubiquinone biosynthesis in maize.

3.3.3. Maize HCT gene repertoire - entry point to the G and S monolignol pathway

HCT enzymes serve as the entry point to the synthesis of G and S lignin subunits. Mutations of this function have not been discovered by classical mutagenesis in maize. A search of the maize genome and the literature indicates that there are at least 42 members of this family in maize (named *ZmHCT1* thru *42*, Supp. Table 1) [65]. We performed a multiple sequence analysis of the 42 predicted maize HCT proteins and

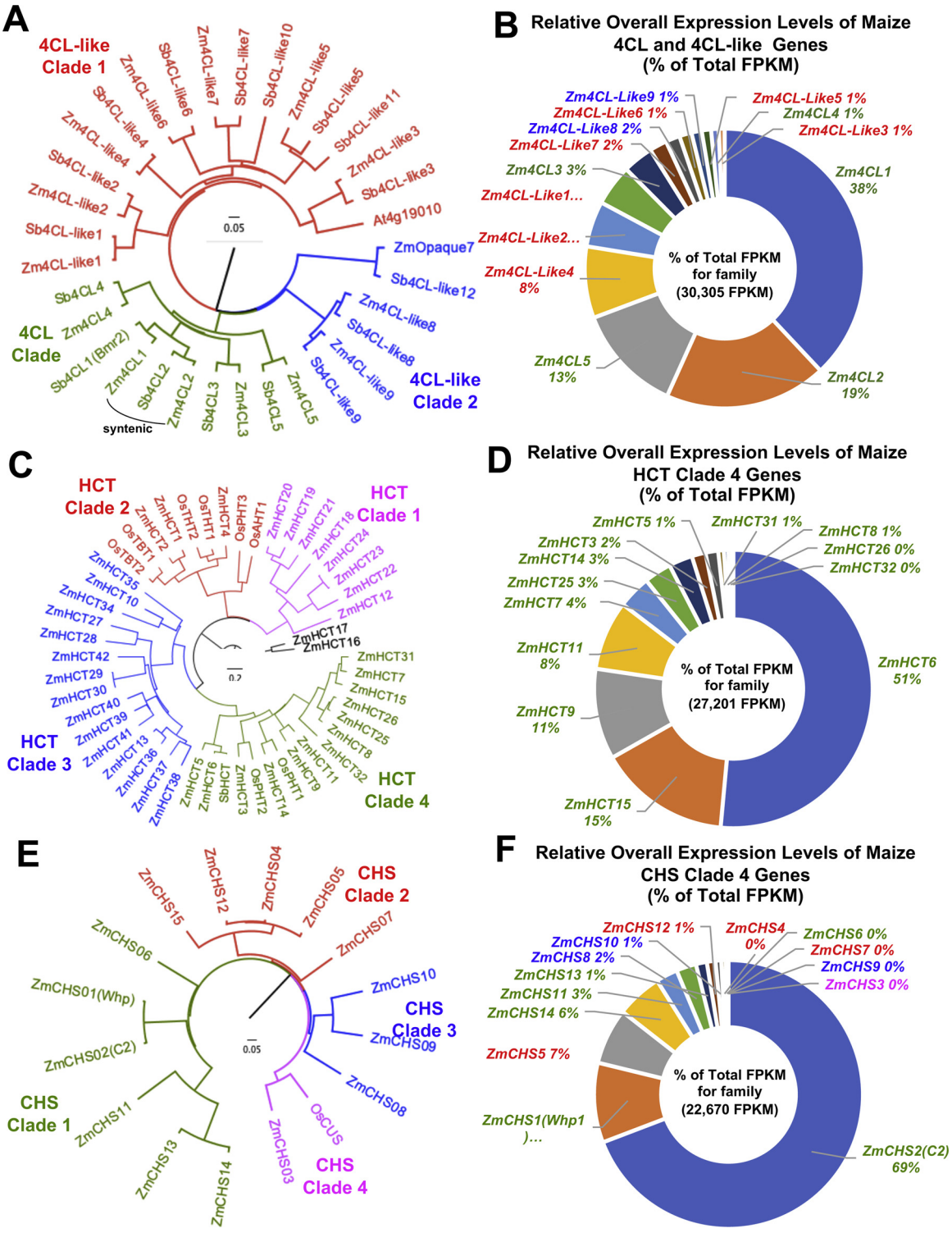

Fig. 6. Phylogenetic Relationships, and Expression Analysis of Maize *4CL*, and *4CL-like*, *HCT*, and *CHS* genes. A, C, and E: Phylogenetic relationship amongst A: maize and sorghum *4CL* and *4CL-like* genes, B: maize *HCT* genes and E: maize *CHS* genes. **B, D, and F :** Pie charts of relative overall expression levels of maize B: *4CL* and *4CL-like* genes, D: *HCT* Clade 4 genes, and F: CHS genes, expressed as a % of total FPKM for the family across 95 tissues (dataset from Ref. [70]).

included also eight characterized BAHD acyl-transferases from rice and one from sorghum. Phylogenetic analysis identified four main *HCT* clades (Fig. 6C), which may reflect shared functionalities amongst clade members. For example, HCT Clade 4 contains SbHCT (SORBI_3004G212300) from sorghum whose crystal structure has been determined [87]. Residues T36, S38, Y40, H162, R371, and T384 play key roles in catalysis and specificity towards *p*-coumaroyl shikimate [87]. We found by multiple sequence alignment (MSA) and structural modeling that these six residues are conserved only in *ZmHCT5* and *6* (data not shown), suggesting that these proteins are responsible for the formation of *p*-coumaroyl shikimate in the early stages of the phenyl-propanoid pathway. *ZmHCT6* is predominantly expressed across 95 maize tissues (51 % of total FPKM for HCT-Clade 4), but *ZmHCT5* is expressed on average at a 34 fold lower rate (1.5 % of total FPKM for

HCT-Clade 4) (Fig. 6D). Other members within the *ZmHCT*-clade4 that are expressed at intermediate levels are *ZmHCT15*, *9*, and *11* (Figs. 6D and 3).

Maize produces a variety of antimicrobial metabolites (phytoalexins) in response to phytopathogen infection. Amongst these are several phenylamides (PA, amine-conjugated phenolic compounds) such as *N-trans*-cinnamoyltryptamine, *N-p*-coumaroylserotonin and *N*-cinnamoyltyramine. Condensation of the hydroxycinnamoyl and amine moieties is the key step for the biosynthesis of PAs and is catalyzed by BAHD acyltransferases [88]. Eight BAHD acyltransferases from rice were also included in the MSA, under the premise of their potential conservation in maize [89,90]. In HCT Clade 2, *ZmHCT1/2* were most similar to rice *OsTBT1/2* which possess tryptamine benzoyl transferase activity [90]. Similarly, ZmHCT4 (in *ZmHCT*-clade 2) is closely related to *OsTHT1* and *2* that exhibit tryptamine hydroxycinnamoyl transferase activity [90]. In HCT Clade 1, *ZmHCT14* and *3* were most similar to *OsPHT1* and *2* respectively which possess putrescine hydroxycinnamoyl transferase activity [91]. Thus, candidate HCT functions in maize could be predicted based on their homology to well characterized acyltransferases in other species. In summary, we found that *ZmHCT*-Clade 4 harbors 4 genes (*ZmHCT6, 15, 9* and *11)* likely to contribute mostly to the flux of metabolites into the G and S monolignol pathway.

3.3.4. Maize CHS gene repertoire - entry point to the flavonoid pathway

The flow of *p*-coumaroyl-coA into the flavonoid pathway is facilitated by CHS and CHI enzymes that are thought to interact via weak but ordered protein-protein interactions [92,93]. Our survey of the maize v4 genome revealed 15 CHS-related genes that could be divided into four minor clades (Fig. 6E). We also included an extra CHS gene (*ZmCHS15*) that was not reported in a recent survey of *CHS* genes in the v3 of the maize genome and our expression analysis employed a much larger set of tissue samples [94]. The classically duplicate genes *ZmCHS1* (*Whp1*) and *ZmCHS2* (*C2*) were found to account respectively for 70 % and 10 % of the overall expression of this gene family (Fig. 6F). Within the same clade they resemble the duplicate genes *ZmCHS13* and *14* on chr 5 and 1 respectively, but the latter are expressed at much lower levels (Fig. 6F). Interestingly the expression of these genes occurs almost exclusively in meiotic tassel tissue along with the homolog *ZmCHS11* (Fig. 4). Notably, these five *CHS* genes collectively account for 89 % of the expression from this gene family. *ZmCHS5* on chr 3 from *ZmCHS*-clade2, appears to function mainly in roots and in drought-stressed roots (Fig. 4) and is another significant contributor (7 % of CHS family FPKMs) (Fig. 6F). CHS genes with *ZmCHS* clades 3 and 4 also appear to play a minor role, but of these, it is of interest that ZmCHS3 is most homologous to OsCUS which catalyzes the formation of bisdemethoxycurcumin, by the condensation of two molecules of 4-coumaroyl-CoA and one molecule of malonyl-CoA [95,96]. So far, this activity has not been reported in maize, but we also observed that *ZmCHS3* expression was undetectable in 95 maize tissues, suggesting that this is a non-functional member of the CHS family in maize. In summary, phylogenetic analyses reveal at least 15 *CHS* and *CHS-like* genes in maize of which only two (*ZmCHS1* and *2*) contribute to 80 % of activity and half of the family members express extremely low or no expression suggesting loss of function following the WGD event.

Overall, our analysis reveals considerable gene subfunctionalization and loss during the evolution of these gene families. Reduced function of half of the *ZmPAL* family as well as a conserved number of *4CL* clade members may have been required to maintain a balanced flux of intermediates through the core pathway since the most recent genome WGD event. The existence of *4CL-like* genes encoding peroxisomal targeted proteins however suggests that a branching of the pathway to ubiquinone biosynthesis may also be conserved in monocots [85,86]. Subfunctionalization may be occurring within different clades of the relatively large HCT family beyond functioning in monolignol production. Some HCT family members may function in phenylamide production and defense functions. Interestingly, *ZmHCT1/2* have been shown to be overexpressed in an autoactive hypersensitive response (HR) mutant of maize [97,98]. Furthermore, evidence indicates that these proteins directly interact with the autoactive Rp1-D21 NLR receptor and caffeoyl CoA O-methyltransferase (CCoAOMT) [97,98] suggesting their involvement in defense related phenolic compound production. Likewise, the Arabidopsis *AtHHT1* (*At5g41040*) gene encodes a feruloyl transferase responsible for synthesizing suberin, is most similar to *ZmHCT10* and *34*, but this activity has not been reported in maize [99]. Subfunctionalization may be occurring also within the *ZmCHS* family, where the *ZmCHS1/2*-homolgous genes *ZmCHS11, 13,* and *14* are expressed at significant levels but almost exclusively in meiotic tassels. These genes were previously reported to increase transiently in response to salicylic acid treatment [94]. Interestingly, they also belong to a co-expression module from meiotic tassels detected by weighted gene co-expression network (WGCNA) analysis [70]. Since flavonoid compounds are undetectable in *ZmCHS1Whp1)* and *ZmCHS2(C2)* mutants [100–102], these polyketide synthases may have evolved to produce a novel metabolite needed for pollen development [103]. Lastly, the fact that a complete *ZmCHS3* gene, which was found to be homologous to *OsCUS*, was not expressed in any of the 95 tissues examined suggests that the ability to synthesize bisdemethoxycurcumin has been lost since the divergence of these species. In summary, the examination of gene families that govern important entry and branch points in the phenylpropanoid pathway allows us to identify those few family members that provide the majority of functional enzymes and those that have a reduced role or have become non-functional. We can also identify family members that have evolved to function either in broader phenolic metabolism or beyond.

3.4. Identification of co-expressed pathway assemblies

Our study identified a large pool of genes potentially associated with phenolic metabolism in maize. To gain further insight into their possible participation in enzyme assemblies, we hypothesized that the genes encoding the components of such enzyme assemblies may also be co-ordinately regulated. Conversely, such assemblies may in turn be predicted by detecting which sequential genes in a pathway are co-expressed. Briefly, we first calculated the mutual rank of the mutual information (MR-MI) of all genes known or predicted to be linked to the biosynthesis of lignin (monolignols-H/S/G), tricin, and flavonoids versus all maize genes using the expression data from 95 maize tissues [104]. This information was then used to define highly co-expressed genes for each phenolic-family member gene (for details see methods). We hypothesized that enzymes in sequential steps of known metabolic pathways are likely to be co-expressed and this was measured by the significance of the connections (*i.e.* those MR values in the top 5 % of co-expressed genes genome-wide) between sequential genes in the pathway (nodes) (Fig. 7 and Supp. Fig. 1). For example, in the core phenylpropanoid pathway, all PALs appear before C4Hs, and 4CLs appear after C4Hs. Within the set of highly co-expressed genes, *ZmPAL1* was highly co-expressed with *ZmC4H1* (MR value = 47) and co-expressed at an intermediate level with *ZmC4H3* (MR value ~ 614), so two connections with different coloration are displayed (Fig. 7A).

For the core phenylpropanoid pathway leading to S monolignol production, 10/10 PALs and 4/5 C4Hs exhibited at least one significant connection but of these *ZmPAL9* shows the highest degree of coordinate expression with *ZmC4H1* and *2* respectively (Fig. 7A). *ZmPAL6/8/*9 and *ZmC4H1/3* were the highest connected nodes (3/3/3 and 16/15 connections, respectively) possibly indicating a major metabolic role in the five pathways that we examined (Fig. 7 and Supp. Fig. 1). In the third step of the pathway, we found that 13/14 of the *4CL/4CL-like* enzyme family had at least one significant connection (Fig. 7A and Supp. Fig. 1A), but of these, *Zm4CL1* shows a high degree of coordinate expression in the S, and G monolignol pathways (Fig. 7A and Supp. Fig. 1B) where this enzyme facilitates the entry of caffeic and ferulic

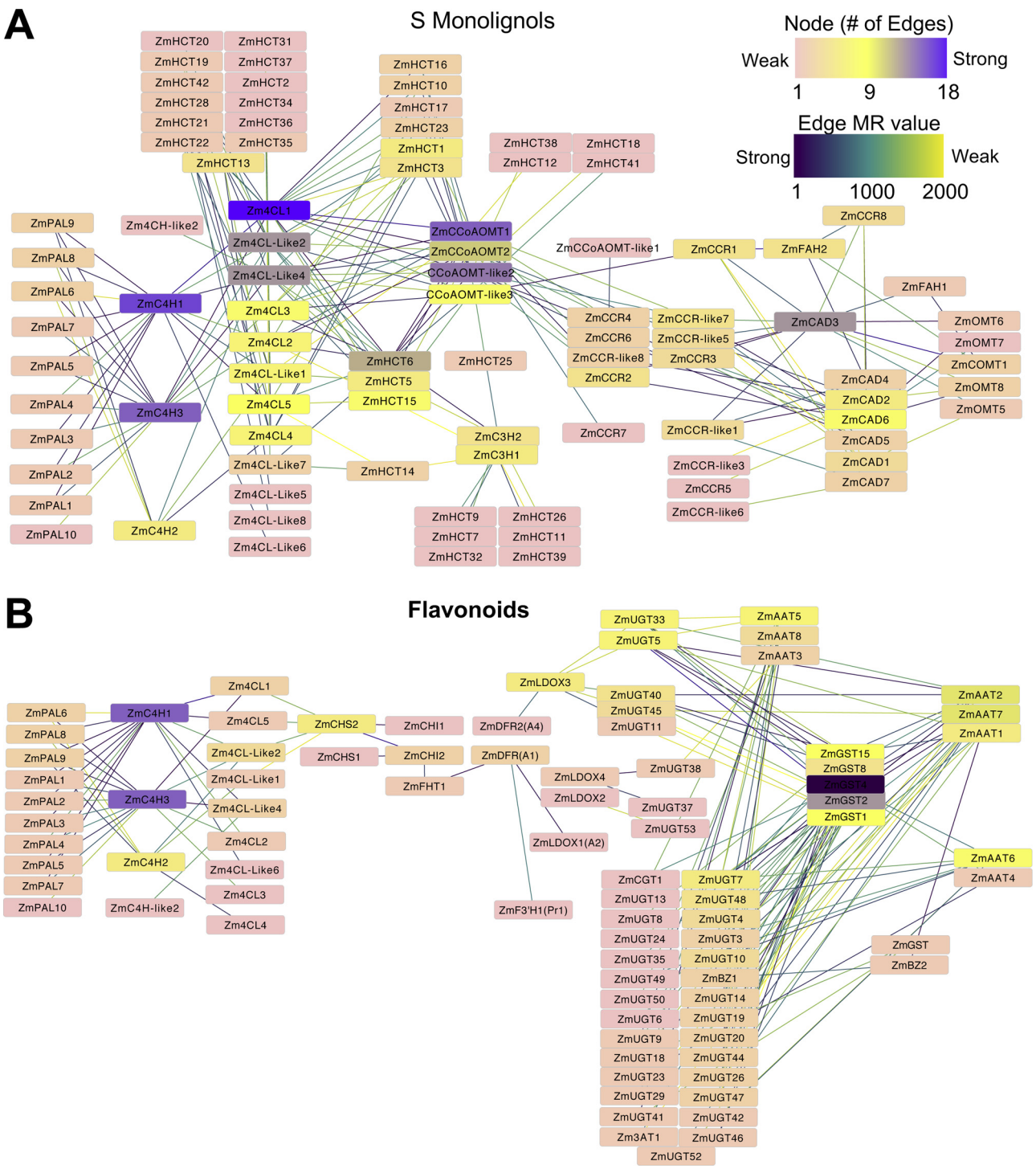

Fig. 7. Co-expression Networks of Core Phenylpropanoid, Monolignol, and Flavonoid Related Genes in Maize. A and B: Diagrams represent the overlap of metabolic pathway and co-expression data associated with the biosynthesis of **A:** core phenylpropanoid molecules and monolignols of type S, and **B:** core phenylpropanoid molecules and anthocyanins respectively. Node color represents the degree of coordinate expression calculated as the number of significant connections (in top 5 % of co-expressed genes) between the sequential genes. Edge color represents the MR value between the sequential genes in the pathway. Family members with no significant links were discarded from the network representation.

acid into the monolignol pathway (Fig. 1). Likewise, *ZmCCoAOMT1, 2* and *ZmCCoAOMT-like2,3,* which operate only in the S and G pathways, are highly connected (Fig. 7A and Supp. Fig. 1B). In the later steps of the monolignol pathways, *ZmCAD3* shows connectivity with the most genes that encode neighboring enzymatic functions in all three pathways (S, G, and H) indicating a major role in the production of all three monolignols (Fig. 7A and Supp. Fig. 1A, B). *ZmCOMT1* and *ZmOMT8* which are amongst the most highly expressed phenolic genes (Fig. 2), exhibit the highest connections amongst all OMTs (Fig. 7A). Noteworthy, for *ZmCOMT1,* these co-expressed genes included the two annotated F5Hs (*ZmFAH1,2)* and *ZmCAD3,* indicating a major role for these enzymes in the production of sinapaldehyde or/and sinapyl alcohol (Fig. 7A).

Hierarchial clustering analysis of core phenylpropanoid and

monolignol related genes also revealed co-expression of genes that may belong to different regulatory modules (Supp. Fig. 2) and this alternative approach permits the clustering of genes whose products do not necessarily operate in a sequential manner. A regulatory module was observed comprising *ZmPAL9 and 1*, *ZmC4H1* and *2*, *ZmCCoAOMT 1 and 2*, and *ZmCOMT1* (Supp. Fig. 2 cluster A). In this module the members are all predominantly expressed, and this module is potentially linked to G or S monolignol production. A nearby cluster comprising *Zm4CL1, ZmCAD3*, and *ZmHCT6* (Supp. Fig. 2 cluster B), may also be linked to G or S monolignol production and of these *Zm4CL1*, and *ZmHCT6* had exhibited high connectivity using MI/MR analysis. A third module comprising *ZmCCoAOMT-like2, ZmCCR-like7*, and *ZmCAD6* (Supp. Fig. 2 cluster C) co-expressed almost exclusively in roots and in roots under drought stress suggesting an enzyme module for monolignol production that is regulated in response to low water stress.

In addition to these clusters, another noteworthy pattern was that certain groups of *HCT* genes co-expressed with genes encoding preceding or subsequent enzymatic functions, or both (Fig. 7 and Supp. Fig. 1). For example, *ZmHCT9,7,11,28,32*, and *39* co-express only with *ZmC3Hs* whereas the *ZmHCT1,3,10,17,18* and *23* co-express with both Zm4CLs and *ZmCCoAOMTs* (Fig. 7A). Although the substrates of most of these HCT enzymes remains unknown, these patterns allow candidate functions for these enzymes to be proposed. For example *ZmHCT6* which is highly expressed (Fig. 3), and co-expressed with *4Cl* and *CCoAOMT* (Fig. 7A), likely encodes the main enzyme responsible for *p*-coumaroyl shikimate production.

Co-expression of genes was also observed in the anthocyanin pathway. The predominantly expressed *ZmCHS1* and *2* were co-expressed with at least one other phenylpropanoid related-gene. Notably, *ZmCHS2 (C2)* exhibited the highest number of connections with *Zm4CL1, 5* and *Zm4CL-Like1* and *2*, and with *ZmCHI1* and *2* (Fig. 7B). We did not identify co-expression links between the F3Hs and F3′Hs that function to metabolize naringenin, however, *ZmDFR (A1)* did co-express with *ZmFHT1* and *ZmF3′H1* (*Pr1*). In the production of cyanidin derivatives, we did identify at least three groups of UGTs that are co-expressed 1) only with LDOXs genes, 2) with GSTs and AATs, and 3) only with AATs (Fig. 7B). This suggests that enzymes that are involved in the final biosynthetic steps may be coordinately regulated and form assemblies to ensure modification and stabilization of these phenolics. Overall, we found that 25/38 of the top 38 predominantly expressed phenylpropanoid genes are also amongst those found to be co-expressed. This suggests that the predominantly expressed genes are also those that are the main participants in forming enzyme assemblies.

Hierarchial clustering analysis of core phenylpropanoid and flavonoid related genes also revealed co-expression of genes that may belong to distinct regulatory modules (Supp. Fig. 3). The largest of these modules consists of 64 gene models that are expressed at low levels across most tissues (Supp. Fig. 3 cluster D). This module indicates that a large number of flavonoid related genes are co-ordinately regulated to provide required metabolites at a low level throughout development. Visual inspection of this large cluster suggests that some sub-clusters are likely regulated in a more tissue specific fashion (e.g. *ZmUGT11, 15, 16, 51 and 54*, and *ZmCHS1 Whp1)*, that appear to be expressed manly in leaves and biotic and abiotic stressed tissues. A cluster comprising *ZmCHS2 (C2)* and *ZmCHI2* (Supp. Fig. 3 cluster E), indicates that the entry steps into the flavonid pathway may form an enzyme complex in maize as has been found in other plant species [93,105]. Cluster E lies amongst the wider group of 19 phenylpropanoid genes that are predominantly expressed throughout most tissues (Fig. 2 and Supp. Fig. 3), indicating that these form the main modules that govern flux through the pathway. In additon to these main modules, it was also possible to detect regulatory modules that appear restricted to more specific cell types. For example, *ZmCHS6*, *ZmUGT43* and *34*, are expressed at low levels but in older leaves and in response to infection (Supp. Fig. 3 cluster F). Genes in a module comprising *Zm4CL-like9* and *ZmCHS11*, *13*, and *14* (Supp. Fig. 3 cluster G) were expressed almost exclusively in meiotic tassel tissue, suggesting a module specific to pollen development. Genes in a co-expression module comprised of *ZmBz2*, *ZmLDOX1 (A2)*, and *ZmAAT1* (Supp. Fig. 3 cluster H) appear to be expressed mainly in brace roots, tassels, anthers and silks, early seed development, drought stressed roots and *C. gramincola* infected leaves (Supp. Fig. 3). This cluster of genes includes classical genes linked to anthocyanin production in maize and the regulatory module may be stress related. Previous studies have shown that anthocyanin content in B73 brace roots has been linked to whitefly infestation and appears to be salicylic acid dependent [106,107].

In summary, we found examples of maize genes that operate sequentially in phenolic pathway branches are indeed coordinately expressed, suggesting that the products of these genes operate in protein complexes or are required in specific cell types or stressed cells [93]. The concept of metabolons proposes that enzyme complexes or hubs form, particularly at branch points in metabolic pathways, and that the distribution of flux within this system may be regulated by the direct competition between enzymes within complexes [92]. Indeed there is evidence from *Arabidopsis* and soybean of a flavonoid metabolon involving CHS and chalcone reductase (CHR) [93,105]. The high level of connection between *ZmC4H1* and *3*, *Zm4Cl1,* and *Zm4CL-like 1* and *2*, suggests that they serve a similar role in maize. Likewise, *ZmCCoAOMT1* and *2* and *ZmCCoAOMT-like2* and *3* may act as a hub in the midpoint of the G and S monolignol pathway, whereas *ZmCAD3* may act as a hub late in the G monolignol branch. Interestingly, in both the G and S monolignol pathways, many significant connections were seen between HCT family members and either the *4CL* or *4CL-like* hubs or the *ZmCCoAOMT1* or *ZmCCoAOMT-like* hubs. This suggests that the substrates are passed from these hubs to HCTs but the specific activities of these HCTs is mostly unknown although in some cases candidate functions may be predicted.

Overall, we find that our approach to detect co-expression modules by MI/MR analysis provides a useful method to identify candidate enzyme assemblies. The display of edge strengths and number of connections between sequential steps provides a facile method to visualize strong and weak candidates for further experimentation. The prominent members of the co-expression modules were also largely in agreement with those found in clusters detected by hierarchial clustering using the same large expression dataset [70]. We also compared our co-expression modules with the 12 modules detected by weighted gene co-expression network analysis (WGCNA) as described in the original analysis of the expression dataset [70]. We found that the genes identified in co-expression modules in this study were disconnected and mainly distributed across three WGCNA modules which contained between 2374 (Module 3) and 4836 (Module 6) members in each module. Therefore, we find that our MI/MR analytical approach permits a more refined and focused approach to predicting gene isoforms whose products may interact in sequential steps of a pathway.

3.5. Assigning candidate regulators to pathway modules

Our analysis revealed patterns of co-expression of several phenylpropanoid genes in the maize genome which logically leads to the hypothesis that they are regulated by the same or overlapping sets of TFs [56,108]. To test this hypothesis, we asked if protein-DNA interactions (PDIs) previously identified for maize TF and 50 phenolic gene targets using a Y1H screen [61] might be co-expressed, and thus lend evidence to their participation in shared regulatory modules. Of the 337 PDIs involving TFs with multiple targets, 21 (6.2 %) were co-expressed and these were between nine TFs and 14 target genes (Fig. 8A). For ZmCPP8, ZmMYB40, ZmMYB109, ZmOrphan249, ZmHSF24 and bZIP76 only one of 4–12 respective Y1H targets was co-expressed, suggesting a less important or more tissue-specific regulatory role for these TFs. However, we found that for ZmMYB65, ZmMYB19 and ZmC3H9, 6/13 (46 %), 5/11 (45 %), and 4/6 (67 %) of their targets

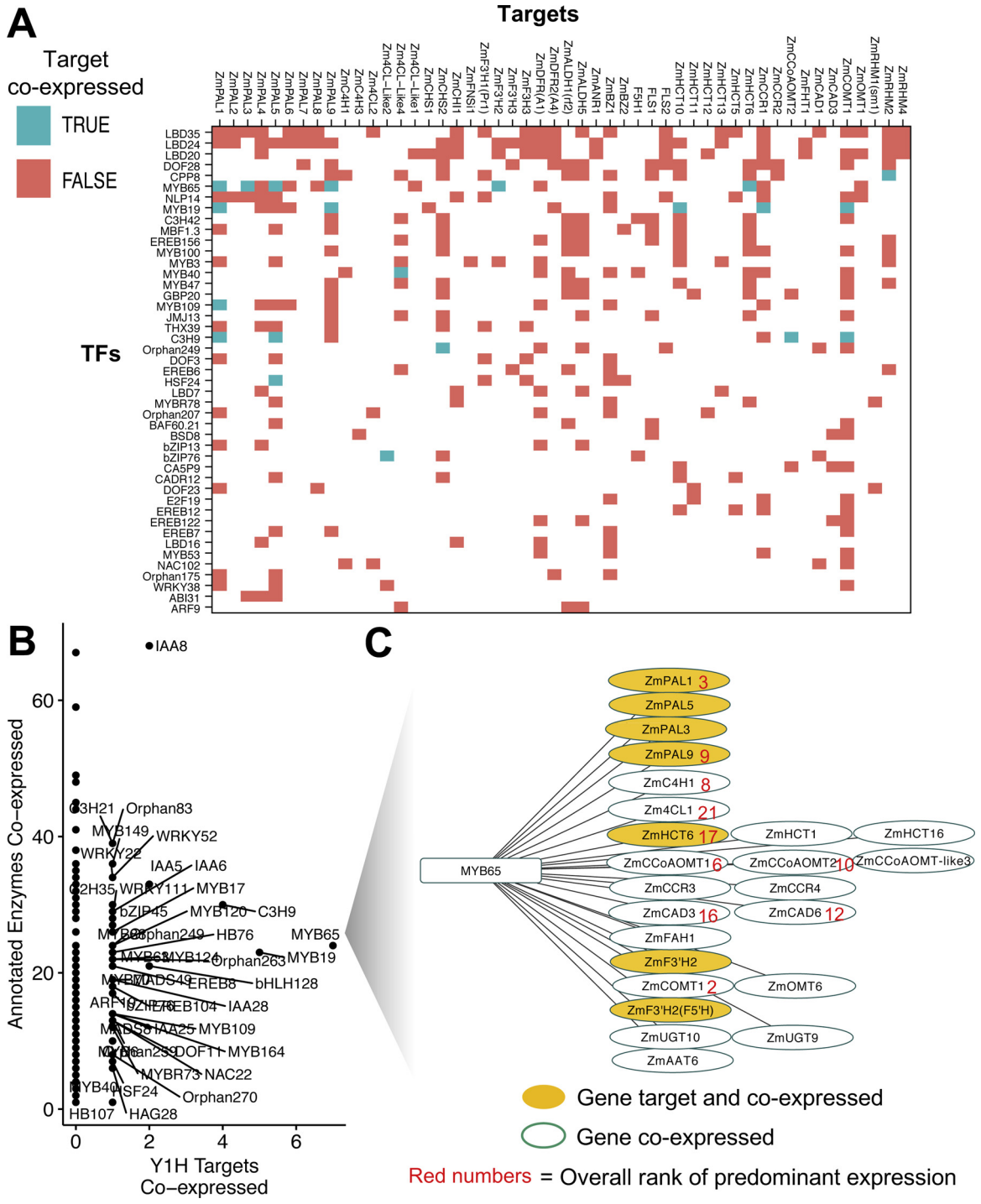

Fig. 8. Candidate Novel Regulatory Links Between Maize TFs and Phenylpropanoid Gene Targets. A: Overlapping of previously discovered Y1H targets (61), and co-expressed TFs with their targets (this study). **B:** Plot of TFs that are co-expressed with a known target(s) defined by Y1H (61) vs the number of other phenylpropanoid targets with which they are co-expressed (this study). **C:** Example of potential new phenylpropanoid targets for MYB65 (white nodes) defined by co-expression in this study. Yellow nodes indicate co-expressed targets that were also defined by Y1H (61). Red numbers indicate the rank of predominantly expressed phenylpropanoid gene (see Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

were co-expressed across the 95 tissues examined, underscoring the likelihood that these are important regulators of these targets *in planta*. Indeed, *ZmMYB19 (Zm1)* and its syntelog in sorghum *SbMYB60* have both been linked to phenylpropanoid regulation [109,110]. The genome-based co-expression approach to finding candidate PDIs permitted us to widen the search for more candidate regulators of phenolic gene expression in maize. We asked how many of the TFs from the Y1H study were co-expressed with the phenolic genes annotated in this study (Fig. 8B and Supp. Tables 1 and 5). We found that 22 TFs with a single Y1H identified target were co-expressed with between 5 and 39 other phenolic genes, for example *ZmC3H21* and *ZmOrphan83* both are co-expressed with 39 phenolic related genes (Fig. 8B and Supp. Table 5). *ZmbHLH128* and *ZmIAA5,* which were found to have 2 targets by Y1H are co-expressed with 21 and 33 phenolic related gene targets respectively (Fig. 8B). ZmMYB65 which was found to have 13 targets by Y1H screen was co-expressed with 24 phenolic genes in this study of which 6

overlapped with those from Y1H (Fig. 8C). Overall, 4430 candidate PDIs were predicted by co-expression analysis (Fig. 8B) with 70 TFs co-expressed with more than 20 genes (*ZmIAA8* is co-expressed with 68 targets, Supp. Table 4). By increasing the co-expression stringency, we reduced this to 804 candidate PDIs (Supp. Table 5). This list included four TFs (*ZmCOL13*, *ZmHB64*, *ZmHB112* and *WRKY16*), with eight candidate targets in the phenylpropanoid pathway and thus they are candidates for master regulators. This more stringent approach permitted us to assign on average four candidate TF regulators to 203 phenylpropanoid genes promoters with 20 promoters having more than eight different candidate TFs (Supp. Table 5).

In summary, we found that many phenylpropanoid genes are co-expressed, suggesting their regulation by similar regulatory modules. By extending the co-expression analysis to include maize TFs, and by using different levels of stringency, we identified 1000's of candidate regulators that were previously unsuspected, and include auxin responsive and homeobox regulators. For some TFs which were co-expressed with a single target identified by Y1H, the number of other phenylpropanoid genes co-expressed was found to be as high as 23 (*i.e.* for *ZmC3H21* and *ZmOrphan83*). Although, the total number of candidate PDIs appears large, it is in the range of 20 TFs per promoter on average. We envisage that this predictive approach will complement Y1H screens by permitting an informed TF-centric screening to be performed with target promoters of interest. By availing of the maize TFome resource, the predicted interactions may quickly be validated by Y1H or transient expression assay so that regulatory modules may be constructed [61,111]. Future advances could involve extending the MI/MR co-expression method to RNA-Seq datasets obtained from single cells thus permitting the highest resolution of metabolons and their regulatory modules.

4. Conclusions

We identified approximately 101 genes from 25 enzyme families in the maize v4 genome, with an average FPKM > 10 across 95 tissues, that make minor to major contributions to phenylpropanoid metabolism. Expression analyses indicate that of these, just 19 genes act predominantly during maize development and response to stress. Related family members have often undergone subfunctionalization, and in some cases phylogenetic comparison permits candidate functions to be predicted. Other genes have undergone pseudogenization, but caution should be applied when evaluating these genes in other germplasm. A MI/MR based co-expression analysis predicts novel co-expression modules that may reflect the formation of metabolons *in planta*. Furthermore, extending this co-expression method to include maize TFs permitted the prediction of numerous PDIs that could contribute to regulatory modules governing phenylpropanoid metabolism in maize. Our comprehensive analysis combined with existing maize TFome resources should permit rapid prediction and evaluation of regulatory modules in maize.

Author contributions

L.G-C compiled and annotated the list of candidate phenylpropanoid genes in *Arabidopsis* and *Arabidopsis* and contributed to writing the manuscript. F.G-C conducted the co-expression and regulatory module analysis. F.D. contributed to gene annotation and writing the manuscript. R.A-V. contributed to the analysis of the PAL gene family. A.D. contributed to funding acquisition, study design, analysis of PAL genes and writing the manuscript. E.G. contributed to funding acquisition, study design, project administration, analysis and writing the manuscript. J.G. conducted expression and phylogenetic analysis, and contributed to funding acquisition, study design, analysis, and writing the manuscript.

Funding sources

Funding for this project was provided by grants National Science Foundation IOS-1733633 to E.G., AID and J.G., by the Agriculture and Food Research Initiative (AFRI) competitive award number 2018-67013-27424 from the USDA National Institute of Food and Agriculture to E.G. and in part by the National Science Foundation Research Traineeship Program (DGE-1828149) to F.A.G.C.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.plantsci.2019.110364.

References

[1] N.G. Lewis, Plant phenolics, Antioxidants in Higher Plants, CRC Press, 2017, pp. 135–169.
[2] L. Chalker-Scott, L.H. Fuchigami, The role of phenolic compounds in plant stress responses, Low Temperature Stress Physiology in Crops, CRC Press, 2018, pp. 67–80.
[3] C.J. Douglas, Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees, Trends Plant Sci. 1 (1996) 171–178.
[4] B. Weisshaar, G.I. Jenkins, Phenylpropanoid biosynthesis and its regulation, Curr. Opin. Plant Biol. 1 (1998) 251–257.
[5] T. Vogt, Phenylpropanoid biosynthesis, Mol. Plant 3 (2010) 2–20.
[6] R. Vanholme, B. De Meester, J. Ralph, W. Boerjan, Lignin biosynthesis and its integration into metabolism, Curr. Opin. Biotechnol. 56 (2019) 230–239.
[7] R. Vanholme, B. Demedts, K. Morreel, J. Ralph, W. Boerjan, Lignin biosynthesis and structure, Plant Physiol. 153 (2010) 895–905.
[8] P. Wang, N. Dudareva, J.A. Morgan, C. Chapple, Genetic manipulation of lignocellulosic biomass for bioenergy, Curr. Opin. Chem. Biol. 29 (2015) 32–39.
[9] F. Muro-Villanueva, X. Mao, C. Chapple, Linking phenylpropanoid metabolism, lignin deposition, and plant growth inhibition, Curr. Opin. Biotechnol. 56 (2019) 202–208.
[10] B. Vanholme, I. El Houari, W. Boerjan, Bioactivity: phenylpropanoids' best kept secret, Curr. Opin. Biotechnol. 56 (2019) 156–162.
[11] E. Grotewold, The genetics and biochemistry of floral pigments, Annu. Rev. Plant Biol. 57 (2006) 761–780.
[12] B. Winkel-Shirley, Flavonoids in seeds and grains: physiological function, agronomic importance and the genetics of biosynthesis, Seed Sci. Res. 8 (1998) 415–422.
[13] B. Winkel-Shirley, Biosynthesis of flavonoids and effects of stress, Curr. Opin. Plant Biol. 5 (2002) 218–223.
[14] L.P. Taylor, E. Grotewold, Flavonoids as developmental regulators, Curr. Opin. Plant Biol. 8 (2005) 317–323.
[15] N.K. Peters, J.W. Frost, S.R. Long, A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes, Science 233 (1986) 977–980.
[16] U. Mathesius, Flavonoid functions in plants and their interactions with other organisms, Plants 7 (2018).
[17] J. Barros, J.C. Serrani-Yarce, F. Chen, D. Baxter, B.J. Venables, R.A. Dixon, Role of bifunctional ammonia-lyase in grass cell wall biosynthesis, Nat. Plants 2 (2016) 16050.
[18] H. Renault, M. De Marothy, G. Jonasson, P. Lara, D.R. Nelson, I. Nilsson, F. André, G. Von Heijne, D. Werck-Reichhart, Gene duplication leads to altered membrane topology of a cytochrome P450 enzyme in seed plants, Mol. Biol. Evol. 34 (2017) 2041–2056.
[19] W. Xiong, Z. Wu, Y. Liu, Y. Li, K. Su, Z. Bai, S. Guo, Z. Hu, Z. Zhang, Y. Bao, J. Sun, G. Yang, C. Fu, Mutation of 4-coumarate: coenzyme A ligase 1 gene affects lignin biosynthesis and increases the cell wall digestibility in maize *brown midrib5* mutants, Biotechnol. Biofuels 12 (2019) 82.
[20] J.R. Andersen, I. Zein, G. Wenzel, B. Darnhofer, J. Eder, M. Ouzunova, T. Lubberstedt, Characterization of phenylpropanoid pathway genes within European maize (*Zea mays* L.) inbreds, BMC Plant Biol. 8 (2008) 2.
[21] D. Cukovic, J. Ehlting, J.A. VanZiffle, C.J. Douglas, Structure and evolution of 4-coumarate:coenzyme A ligase (4CL) gene families, Biol. Chem. 382 (2001) 645–654.
[22] L. Hoffmann, S. Maury, F. Martz, P. Geoffroy, M. Legrand, Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism, J. Biol. Chem. 278 (2003) 95–103.
[23] R. Vanholme, I. Cesarino, K. Rataj, Y. Xiao, L. Sundin, G. Goeminne, H. Kim, J. Cross, K. Morreel, P. Araujo, Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis, Science 341 (2013) 1103–1106.

[24] G. Schoch, S. Goepfert, M. Morant, A. Hehn, D. Meyer, P. Ullmann, D. Werck-Reichhart, CYP98A3 from *Arabidopsis thaliana* is a 3′-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway, J. Biol. Chem. 276 (2001) 36566–36574.

[25] A.M. Walker, S.A. Sattler, M. Regner, J.P. Jones, J. Ralph, W. Vermerris, S.E. Sattler, C. Kang, The structure and catalytic mechanism of *Sorghum bicolor* caffeoyl-CoA O-methyltransferase, Plant Physiol. 172 (2016) 78–92.

[26] H.L. Park, S.H. Bhoo, M. Kwon, S.-W. Lee, M.-H. Cho, Biochemical and expression analyses of the rice cinnamoyl-CoA reductase gene family, Front. Plant Sci. 8 (2017) 2099.

[27] S.-Y. Jun, A.M. Walker, H. Kim, J. Ralph, W. Vermerris, S.E. Sattler, C. Kang, The enzyme activity and substrate specificity of two major cinnamyl alcohol dehydrogenases in sorghum (*Sorghum bicolor*), SbCAD2 and SbCAD4, Plant Physiol. 174 (2017) 2128–2145.

[28] K. Meyer, J.C. Cusumano, C. Somerville, C.C. Chapple, Ferulate-5-hydroxylase from *Arabidopsis thaliana* defines a new family of cytochrome P450-dependent monooxygenases, Proc Natl Acad Sci U S A 93 (1996) 6869–6874.

[29] A.R. Green, K.M. Lewis, J.T. Barr, J.P. Jones, F. Lu, J. Ralph, W. Vermerris, S.E. Sattler, C. Kang, Determination of the structure and catalytic mechanism of *Sorghum bicolor* caffeic acid O-methyltransferase and the structural impact of three brown midrib12 mutations, Plant Physiol. 165 (2014) 1440–1456.

[30] Y. Tobimatsu, M. Schuetz, Lignin polymerization: how do plants manage the chemistry so well? Curr. Opin. Biotechnol. 56 (2019) 75–81.

[31] J.L. Ferrer, J.M. Jez, M.E. Bowman, R.A. Dixon, J.P. Noel, Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis, Nat. Struct. Biol. 6 (1999) 775–784.

[32] E. Grotewold, T. Peterson, Isolation and characterization of a maize gene encoding chalcone flavonone isomerase, Mol. Gen. Genet. 242 (1994) 1–8.

[33] F. Brugliera, G. Barri-Rewell, T.A. Holton, J.G. Mason, Isolation and characterization of a flavonoid 3'-hydroxylase cDNA clone corresponding to the Ht1 locus of Petunia hybrida, Plant J. 19 (1999) 441–451.

[34] M.L. Falcone Ferreyra, J. Emiliani, E.J. Rodriguez, V.A. Campos-Bermudez, E. Grotewold, P. Casati, The identification of maize and Arabidopsis type I FLAVONE SYNTHASEs links flavones with hormones and biotic interactions, Plant Physiol. 169 (2015) 1090–1107.

[35] L. Britsch, W. Heller, H. Grisebach, Conversion of flavanone to flavone, dihydroflavonol and flavonol with an enzyme system from cell cultures of parsley, Zeitschrift fur Naturforschung - Section C Journal of Biosciences 36 (1981) 742–750.

[36] S. Martens, G. Forkmann, Cloning and expression of flavone synthase II from Gerbera hybrids, Plant J. 20 (1999) 611–618.

[37] S. Righini, E.J. Rodriguez, C. Berosich, E. Grotewold, P. Casati, M.L. Falcone Ferreyra, Apigenin produced by maize flavone synthase I and II protects plants against UV-B-induced damage, Plant Cell Environ. 42 (2019) 495–508.

[38] D.K. Owens, K.C. Crosby, J. Runac, B.A. Howard, B.S. Winkel, Biochemical and genetic characterization of Arabidopsis flavanone 3beta-hydroxylase, Plant Physiol. Biochem. 46 (2008) 833–843.

[39] G.B. Deboo, M.C. Albertsen, L.P. Taylor, Flavanone 3-hydroxylase transcripts and flavonol accumulation are temporally coordinated in maize anthers, Plant J. 7 (1995) 703–713.

[40] Y. Mo, C. Nagel, L.P. Taylor, Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 7213–7217.

[41] M.L. Falcone Ferreyra, S. Rius, J. Emiliani, L. Pourcel, A. Feller, K. Morohashi, P. Casati, E. Grotewold, Cloning and characterization of a UV-B-inducible maize flavonol synthase, Plant J. 62 (2010) 77–91.

[42] S. Martens, A. Preuss, U. Matern, Multifunctional flavonoid dioxygenases: flavonol and anthocyanin biosynthesis in Arabidopsis thaliana L, Phytochemistry 71 (2010) 1040–1049.

[43] A.R. Reddy, L. Britsch, F. Salamini, H. Saedler, W. Rohde, The *A1* (anthocyanin-1) locus in *Zea mays* encodes dihydroquercitin reductase, Plant Sci. 52 (1987) 7–13.

[44] J. Bernhardt, K. Stich, Z. Schwarz-Sommer, H. Saedler, U. Wienand, Molecular analysis of a second functional *A1* gene (dihydroflavonol 4-reductase) in *Zea mays*, Plant J. 14 (1998) 483–488.

[45] E. Grotewold, B.J. Drummond, B. Bowen, T. Peterson, The myb-homologous *P* gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset, Cell 76 (1994) 543–553.

[46] M.K. Pelletier, J.R. Murrell, B.W. Shirley, Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in Arabidopsis. Further evidence for differential regulation of "early" and "late" genes, Plant Physiol. 113 (1997) 1437–1445.

[47] A. Menssen, S. Hohmann, W. Martin, P.S. Schnable, P.A. Peterson, H. Saedler, A. Gierl, The En/Spm transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a *dSpm* element in the *A2* gene, EMBO J. 9 (1990) 3051–3057.

[48] G.M. Reddy, Genetic control of leucoanthocyanidin formation in maize, Genetics 50 (1964) 485–489.

[49] M. Gargouri, C. Manigand, C. Mauge, T. Granier, B. Langlois d'Estaintot, O. Cala, I. Pianet, K. Bathany, J. Chaudiere, B. Gallois, Structure and epimerase activity of anthocyanidin reductase from *Vitis vinifera*, Acta Crystallogr. D Biol. Crystallogr. 65 (2009) 989–1000.

[50] J. Zhao, Flavonoid transport mechanisms: how to go, and with whom, Trends Plant Sci. 20 (2015) 576–585.

[51] D. Bowles, A multigene family of glycosyltransferases in a model plant, Arabidopsis thaliana, Biochem. Soc. Trans. 30 (2002) 301–306.

[52] H. Suzuki, T. Nakayama, K. Yonekura-Sakakibara, Y. Fukui, N. Nakamura, M.A. Yamaguchi, Y. Tanaka, T. Kusumi, T. Nishino, cDNA cloning, heterologous expressions, and functional characterization of malonyl-coenzyme a:anthocyanidin 3-O-glucoside-6"-O-malonyltransferase from dahlia flowers, Plant Physiol. 130 (2002) 2142–2151.

[53] M.R. Alfenito, E. Souer, C.D. Goodman, R. Buell, J. Mol, R. Koes, V. Walbot, Functional complementation of anthocyanin sequestration in the vacuole by widely divergent glutathione S-transferases, Plant Cell 10 (1998) 1135–1149.

[54] M.I. Casas, M.L. Falcone-Ferreyra, N. Jiang, M.K. Mejia-Guerra, E. Rodriguez, T. Wilson, J. Engelmeier, P. Casati, E. Grotewold, Identification and characterization of maize salmon silks genes involved in insecticidal maysin biosynthesis, Plant Cell 28 (2016) 1297–1309.

[55] E.H. Coe Jr., The origins of maize genetics, Nature reviews 2 (2001) 898–905.

[56] J. Gray, D. Caparros-Ruiz, E. Grotewold, Grass phenylpropanoids: regulate before using!, Plant Sci. 184 (2012) 112–120.

[57] J. Liu, A. Osbourn, P. Ma, MYB transcription factors as regulators of phenylpropanoid metabolism in plants, Mol. Plant 8 (2015) 689–708.

[58] M. Ohtani, T. Demura, The quest for transcriptional hubs of lignin biosynthesis: beyond the NAC-MYB-gene regulatory network model, Curr. Opin. Biotechnol. 56 (2019) 82–87.

[59] D. Ma, C.P. Constabel, MYB repressors as regulators of phenylpropanoid metabolism in plants, Trends Plant Sci. 24 (2019) 275–289.

[60] B. Burdo, J. Gray, M.P. Goetting-Minesky, B. Wittler, M. Hunt, T. Li, D. Velliquette, J. Thomas, I. Gentzel, M.D. Brito, M.K. Mejia-Guerra, L.N. Connolly, D. Qaisi, W. Li, M.I. Casas, A.I. Doseff, E. Grotewold, The Maize TFome - development of a transcription factor open reading frame collection for functional genomics, Plant J. 80 (2014) 356–366.

[61] F. Yang, W. Li, N. Jiang, H. Yu, K. Morohashi, W.Z. Ouma, D.E. Morales-Mantilla, F.A. Gomez-Cano, E. Mukundi, L.D. Prada-Salcedo, R.A. Velazquez, J. Valentin, M.K. Mejia-Guerra, J. Gray, A.I. Doseff, E. Grotewold, A maize gene regulatory network for phenolic metabolism, Mol. Plant 10 (2017) 498–515.

[62] M.A. Costa, R.E. Collins, A.M. Anterola, F.C. Cochrane, L.B. Davin, N.G. Lewis, An *in silico* assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in *Arabidopsis thaliana* and limitations thereof, Phytochemistry 64 (2003) 1097–1112.

[63] J. Ehlting, N. Mattheus, D.S. Aeschliman, E. Li, B. Hamberger, I.F. Cullis, J. Zhuang, M. Kaneda, S.D. Mansfield, L. Samuels, K. Ritland, B.E. Ellis, J. Bohlmann, C.J. Douglas, Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation, Plant J. 42 (2005) 618–640.

[64] B. Hamberger, M. Ellis, M. Friedmann, C. de Azevedo Souza, B. Barbazuk, C.J. Douglas, Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the Populus lignin toolbox and conservation and diversification of angiosperm gene familiesThis article is one of a selection of papers published in the Special Issue on Poplar Research in Canada, Can. J. Bot. 85 (2007) 1182–1201.

[65] B.W. Penning, C.T. Hunter 3rd, R. Tayengwa, A.L. Eveland, C.K. Dugard, A.T. Olek, W. Vermerris, K.E. Koch, D.R. McCarty, M.F. Davis, S.R. Thomas, M.C. McCann, N.C. Carpita, Genetic resources for maize cell wall biology, Plant Physiol. 151 (2009) 1703–1728.

[66] J.C. Schnable, N.M. Springer, M. Freeling, Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss, Proc Natl Acad Sci U S A 108 (2011) 4069–4074.

[67] H. Tang, M.D. Bomhoff, E. Briones, L. Zhang, J.C. Schnable, E. Lyons, SynFind: compiling syntenic regions across any set of genomes on demand, Genome Biol. Evol. 7 (2015) 3286–3298.

[68] A. Rambaut, FigTree v1.4.4, Available from (2014) http://tree.bio.ed.ac.uk/software/.

[69] D. Kim, J.M. Paggi, C. Park, C. Bennett, S.L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, Nat. Biotechnol. 37 (2019) 907–915.

[70] G.M. Hoopes, J.P. Hamilton, J.C. Wood, E. Esteban, A. Pasha, B. Vaillancourt, N.J. Provart, C.R. Buell, An updated gene atlas for maize reveals organ-specific and stress-induced genes, Plant J. 97 (2019) 1154–1167.

[71] F.E.J. Harrell, R Package 'Hmisc', Available from (2019) https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf.

[72] A. Kraskov, H. Stogbauer, P. Grassberger, Estimating mutual information, Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 69 (2004) 066138.

[73] G. Sales, C. Romualdi, Parmigene–a parallel R package for mutual information estimation and gene network reconstruction, Bioinformatics 27 (2011) 1876–1877.

[74] R. Suzuki, R package 'pvclust', Available from (2015) https://cran.r-project.org/web/packages/pvclust/pvclust.pdf.

[75] Z. Zhang, Z. Liang, L. Yin, Q.X. Li, Z. Wu, Distribution of four bioactive flavonoids in maize tissues of five varieties and correlation with expression of the biosynthetic genes, J. Agric. Food Chem. 66 (2018) 10431–10437.

[76] K. Wittmeyer, J. Cui, D. Chatterjee, T.F. Lee, Q. Tan, W. Xue, Y. Jiao, P.H. Wang, I. Gaffoor, D. Ware, B.C. Meyers, S. Chopra, The dominant and poorly penetrant phenotypes of maize unstable factor for orange1 are caused by DNA methylation changes at a linked transposon, Plant Cell 30 (2018) 3006–3023.

[77] G.B. Deboo, M.C. Albertsen, L.P. Taylor, Flavanone 3-hydrosylase transcripts and flavonol accumulation are temporally coordiante in maize anthers, Plant J. 7 (1995) 703–713.

[78] M.N. Paulsmeyer, P.J. Brown, J.A. Juvik, Discovery of anthocyanin acyltransferase1 (AAT1) in maize using genotyping-by-sequencing (GBS), G3: Genes - Genomes - Genetics 8 (2018) 3669.

[79] S.Y. Jun, S.A. Sattler, G.S. Cortez, W. Vermerris, S.E. Sattler, C. Kang, Biochemical and structural analysis of substrate specificity of a phenylalanine ammonia-lyase, Plant Physiol. 176 (2018) 1452–1468.

[80] Z. Wu, S. Gui, S. Wang, Y. Ding, Molecular evolution and functional characterisation of an ancient phenylalanine ammonia-lyase gene (*NnPAL1*) from *Nelumbo nucifera*: novel insight into the evolution of the PAL family in angiosperms, BMC Evol. Biol. 14 (2014) 100.

[81] A. Saballos, S.E. Sattler, E. Sanchez, T.P. Foster, Z. Xin, C. Kang, J.F. Pedersen, W. Vermerris, *Brown midrib2* (*Bmr2*) encodes the major 4-coumarate:coenzyme A ligase involved in lignin biosynthesis in sorghum (*Sorghum bicolor* (L.) Moench), Plant J. 70 (2012) 818–830.

[82] J.M. Shockey, M.S. Fulda, J. Browse, Arabidopsis contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme a synthetases, Plant Physiol. 132 (2003) 1065–1076.

[83] Z. Li, S.K. Nair, Structural basis for specificity and flexibility in a plant 4-Coumarate:CoA ligase, Structure 23 (2015) 2032–2042.

[84] K. Schneider, K. Hovel, K. Witzel, B. Hamberger, D. Schomburg, E. Kombrink, H.P. Stuible, The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase, Proc Natl Acad Sci U S A 100 (2003) 8601–8606.

[85] A. Block, J.R. Widhalm, A. Fatihi, R.E. Cahoon, Y. Wamboldt, C. Elowsky, S.A. Mackenzie, E.B. Cahoon, C. Chapple, N. Dudareva, G.J. Basset, The origin and biosynthesis of the benzenoid moiety of ubiquinone (Coenzyme Q) in Arabidopsis, Plant Cell 26 (2014) 1938–1948.

[86] E. Soubeyrand, T.S. Johnson, S. Latimer, A. Block, J. Kim, T.A. Colquhoun, E. Butelli, C. Martin, M.A. Wilson, G.J. Basset, The peroxidative cleavage of kaempferol contributes to the biosynthesis of the benzenoid moiety of Ubiquinone in plants, Plant Cell 30 (2018) 2910–2921.

[87] A.M. Walker, R.P. Hayes, B. Youn, W. Vermerris, S.E. Sattler, C. Kang, Elucidation of the structure and reaction mechanism of sorghum hydroxycinnamoyltransferase and its structural relationship to other coenzyme a-dependent transferases and synthases, Plant Physiol. 162 (2013) 640–651.

[88] M. Petersen, Hydroxycinnamoyltransferases in plant metabolism, Phytochem. Rev. 15 (2016) 699–727.

[89] M.H. Cho, S.W. Lee, Phenolic phytoalexins in rice: biological functions and biosynthesis, Int. J. Mol. Sci. 16 (2015) 29120–29133.

[90] M. Peng, Y. Gao, W. Chen, W. Wang, S. Shen, J. Shi, C. Wang, Y. Zhang, L. Zou, S. Wang, J. Wan, X. Liu, L. Gong, J. Luo, Evolutionarily distinct BAHD N-acyltransferases are responsible for natural variation of aromatic amine conjugates in rice, Plant Cell 28 (2016) 1533–1550.

[91] H. Peng, R.S. Meyer, T. Yang, B.D. Whitaker, F. Trouth, L. Shangguan, J. Huang, A. Litt, D.P. Little, H. Ke, W.M. Jurick 2nd, A novel hydroxycinnamoyl transferase for synthesis of hydroxycinnamoyl spermine conjugates in plants, BMC Plant Biol. 19 (2019) 261.

[92] K.C. Crosby, A. Pietraszewska-Bogiel, T.W. Gadella Jr., B.S. Winkel, Forster resonance energy transfer demonstrates a flavonoid metabolon in living plant cells that displays competitive interactions between enzymes, FEBS Lett. 585 (2011) 2193–2198.

[93] T. Nakayama, S. Takahashi, T. Waki, Formation of flavonoid metabolons: functional significance of protein-protein interactions and impact on flavonoid chemodiversity, Front. Plant Sci. 10 (2019) 821.

[94] Y. Han, T. Ding, B. Su, H. Jiang, Genome-wide identification, characterization and expression analysis of the chalcone synthase family in maize, Int. J. Mol. Sci. 17 (2016).

[95] Y. Katsuyama, M. Matsuzawa, N. Funa, S. Horinouchi, In vitro synthesis of curcuminoids by type III polyketide synthase from *Oryza sativa*, J. Biol. Chem. 282 (2007) 37702–37709.

[96] H. Morita, K. Wanibuchi, H. Nii, R. Kato, S. Sugio, I. Abe, Structural basis for the one-pot formation of the diarylheptanoid scaffold by curcuminoid synthase from *Oryza sativa*, Proc Natl Acad Sci U S A 107 (2010) 19778–19783.

[97] G.F. Wang, P.J. Balint-Kurti, Maize homologs of CCoAOMT and HCT, two key enzymes in lignin biosynthesis, form complexes with the NLR Rp1 protein to modulate the defense response, Plant Physiol. 171 (2016) 2166–2177.

[98] G.F. Wang, Y. He, R. Strauch, B.A. Olukolu, D. Nielsen, X. Li, P.J. Balint-Kurti, Maize homologs of hydroxycinnamoyltransferase, a key enzyme in lignin biosynthesis, bind the nucleotide binding leucine-rich repeat Rp1 proteins to modulate the defense response, Plant Physiol. 169 (2015) 2230–2243.

[99] I. Molina, Y. Li-Beisson, F. Beisson, J.B. Ohlrogge, M. Pollard, Identification of an Arabidopsis feruloyl-coenzyme A transferase required for suberin synthesis, Plant Physiol. 151 (2009) 1317–1328.

[100] E.H.J. Coe, S.M. McCormick, S.A. Modena, White pollen in maize, J. Hered. 72 (1981) 318–320.

[101] N.B. Eloy, W. Voorend, W. Lan, M.L. Saleme, I. Cesarino, R. Vanholme, R.A. Smith, G. Goeminne, A. Pallidis, K. Morreel, J. Nicomedes Jr., J. Ralph, W. Boerjan, Silencing CHALCONE SYNTHASE in maize impedes the incorporation of tricin into lignin and increases lignin content, Plant Physiol. 173 (2017) 998–1016.

[102] P. Franken, U. Niesbach-Klosgen, U. Weydemann, L. Marechal-Drouard, H. Saedler, U. Wienand, The duplicated chalcone synthase genes C2 and Whp (white pollen) of *Zea mays* are independently regulated; evidence for translational control of Whp expression by the anthocyanin intensifying gene in, EMBO J. 10 (1991) 2605–2612.

[103] G. Yadav, R.S. Gokhale, D. Mohanty, Towards prediction of metabolic products of polyketide synthases: an in silico analysis, PLoS Comput. Biol. 5 (2009) e1000351.

[104] J.W. Walley, R.C. Sartor, Z. Shen, R.J. Schmitz, K.J. Wu, M.A. Urich, J.R. Nery, L.G. Smith, J.C. Schnable, J.R. Ecker, S.P. Briggs, Integration of omic networks in a developmental atlas of maize, Science 353 (2016) 814–818.

[105] B.S. Winkel, Metabolic channeling in plants, Annu. Rev. Plant Biol. 55 (2004) 85–107.

[106] Y.S. Park, D.W. Bae, C.M. Ryu, Aboveground whitefly infestation modulates transcriptional levels of anthocyanin biosynthesis and jasmonic acid signaling-related genes and augments the cope with drought stress of maize, PLoS One 10 (2015) e0143879.

[107] Y. Yan, S. Christensen, T. Isakeit, J. Engelberth, R. Meeley, A. Hayward, R.J. Emery, M.V. Kolomiets, Disruption of OPR7 and OPR8 reveals the versatile functions of jasmonic acid in maize development and defense, Plant Cell 24 (2012) 1420–1436.

[108] H.K. Dooner, Coordinate genetic regulation of flavonoid biosynthetic enzymes in maize, Molecular and General Genetics MGG 189 (1983) 136–141.

[109] P. Franken, S. Schrell, P.A. Peterson, H. Saedler, U. Wienand, Molecular analysis of protein domain function encoded by the myb-homologous maize genes *C1*, *Zm1* and *Zm38*, Plant J. 6 (1994) 21–30.

[110] E.D. Scully, T. Gries, G. Sarath, N.A. Palmer, L. Baird, M.J. Serapiglia, B.S. Dien, A.A. Boateng, Z. Ge, D.L. Funnell-Harris, P. Twigg, T.E. Clemente, S.E. Sattler, Overexpression of SbMyb60 impacts phenylpropanoid biosynthesis and alters secondary cell wall composition in Sorghum bicolor, Plant J. 85 (2016) 378–395.

[111] A. Yilmaz, M.Y. Nishiyama Jr., B.G. Fuentes, G.M. Souza, D. Janies, J. Gray, E. Grotewold, GRASSIUS: a platform for comparative regulatory genomics across the grasses, Plant Physiol. 149 (2009) 171–180.

[112] J. Rosler, F. Krekel, N. Amrhein, J. Schmid, Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity, Plant Physiol. 113 (1997) 175–179.

[113] Y. Barrière, S. Guillaumie, D. Denoue, M. Pichon, D. Goffner, J.-P. Martinant, Investigating the unusually high cell wall digestibility of the old INRA early flint F4 maize inbred line, Maydica 62 (2017).

[114] Y. Reinprecht, G.E. Perry, K. Peter Pauls, A comparison of phenylpropanoid pathway gene families in common bean. Focus on P450 and C4H genes, in: M.P.érez de la Vega, M. Santalla, F. Marsolais (Eds.), The Common Bean Genome, Springer International Publishing, Cham, 2017, pp. 219–261.

[115] A.M. Gulick, Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase, ACS Chem. Biol. 4 (2009) 811–827.

[116] S. Schmelz, J.H. Naismith, Adenylate-forming enzymes, Curr. Opin. Struct. Biol. 19 (2009) 666–671.

[117] H. Pan, R. Zhou, G.V. Louie, J.K. Mühlemann, E.K. Bomati, M.E. Bowman, N. Dudareva, R.A. Dixon, J.P. Noel, X. Wang, Structural studies of cinnamoyl-CoA reductase and cinnamyl-alcohol dehydrogenase, key enzymes of monolignol biosynthesis, Plant Cell 26 (2014) 3709.

[118] R.A. Smith, C.L. Cass, M. Mazaheri, R.S. Sekhon, M. Heckwolf, H. Kaeppler, N. de Leon, S.D. Mansfield, S.M. Kaeppler, J.C. Sedbrook, S.D. Karlen, J. Ralph, Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolignol ferulates incorporated into maize lignins, Biotechnol. Biofuels 10 (2017) 109.

[119] B. Tamasloukht, M.S.-J. Wong Quai Lam, Y. Martinez, K. Tozo, O. Barbier, C. Jourda, A. Jauneau, G. Borderies, S. Balzergue, J.-P. Renou, S. Huguet, J.P. Martinant, C. Tatout, C. Lapierre, Y. Barrière, D. Goffner, M. Pichon, Characterization of a cinnamoyl-CoA reductase 1 (CCR1) mutant in maize: effects on lignification, fibre development, and global gene expression, J. Exp. Bot. 62 (2011) 3837–3848.

[120] W. Chen, N. VanOpdorp, D. Fitzl, J. Tewari, P. Friedemann, T. Greene, S. Thompson, S. Kumpatla, P. Zheng, Transposon insertion in a cinnamyl alcohol dehydrogenase gene is responsible for a *brown midrib1* mutation in maize, Plant Mol. Biol. 80 (2012) 289–297.

[121] C. Halpin, K. Holt, J. Chojecki, D. Oliver, B. Chabbert, B. Monties, K. Edwards, A. Barakate, G.A. Foxon, *Brown-midrib maize* (*bm1*)-a mutation affecting the cinnamyl alcohol dehydrogenase gene, Plant J. 14 (1998) 545–553.

[122] T.A. Morrison, Activity of phenylalanine ammonia-lyase, tyrosine ammonia-lyase, and cinnamyl alcohol dehydrogenase in the maize stalk, Crop Sci. 33 (1266) (1993) 1264-1268-1993 v.1233.

[123] J. Barros, L. Escamilla-Trevino, L. Song, X. Rao, J.C. Serrani-Yarce, M.D. Palacios, N. Engle, F.K. Choudhury, T.J. Tschaplinski, B.J. Venables, R. Mittler, R.A. Dixon, 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase, Nat. Commun. 10 (2019) 1994.

[124] S. Fornale, J. Rencoret, L. Garcia-Calvo, M. Capellades, A. Encina, R. Santiago, J. Rigau, A. Gutierrez, J.C. Del Rio, D. Caparros-Ruiz, Cell wall modifications triggered by the down-regulation of Coumarate 3-hydroxylase-1 in maize, Plant Sci. 236 (2015) 272–282.

[125] X. Li, W. Chen, Y. Zhao, Y. Xiang, H. Jiang, S. Zhu, B. Cheng, Downregulation of caffeoyl-CoA O-methyltransferase (CCoAOMT) by RNA interference leads to reduced lignin production in maize straw, Genet. Mol. Biol. 36 (2013) 540–546.

[126] F. Vignols, J. Rigau, M.A. Torres, M. Capellades, P. Puigdomènech, The *brown midrib3* (*bm3*) mutation in maize occurs in the gene encoding caffeic acid O-methyltransferase, Plant Cell 7 (1995) 407.

[127] J.-M. Zhou, Y. Fukushi, E. Wollenweber, R.K. Ibrahim, Characterization of two O.-methyltransferase-like genes in barley and maize, Pharm. Biol. 46 (2008) 26–34.

[128] S.J. Szalma, M.E. Snook, B.S. Bushman, K.E. Houchins, M.D. McMullen, Duplicate loci as QTL, Crop Sci. 42 (2002) 1679–1687.

[129] M. Gensheimer, A. Mushegian, Chalcone isomerase family and fold: no longer unique to plants, Protein Sci. 13 (2004) 540–544.

[130] R.L. Larson, J.B. Bussard, Microsomal flavonoid 3'-monooxygenase from maize seedlings, Plant Physiol. 80 (1986) 483–486.

[131] M. Sharma, C. Chai, K. Morohashi, E. Grotewold, M.E. Snook, S. Chopra, Expression of flavonoid 3'-hydroxylase is controlled by P1, the regulator of 3-deoxyflavonoid biosynthesis in maize, BMC Plant Biol. 12 (2012) 196.

[132] V. Martinez, T.C. Mestre, F. Rubio, A. Girones-Vilaplana, D.A. Moreno, R. Mittler, R.M. Rivero, Accumulation of flavonols over hydroxycinnamic acids favors

oxidative damage protection under abiotic stress, Front. Plant Sci. 7 (2016).
[133] K. Morohashi, M.I. Casas, M.L. Falcone Ferreyra, M.K. Mejía-Guerra, L. Pourcel, A. Yilmaz, A. Feller, B. Carvalho, J. Emiliani, E. Rodriguez, S. Pellegrinet, M. McMullen, P. Casati, E. Grotewold, A genome-wide regulatory framework identifies maize *pericarp color1* controlled genes, Plant Cell 24 (2012) 2745–2764.
[134] P. Petit, T. Granier, B.L. d'Estaintot, C. Manigand, K. Bathany, J.M. Schmitter, V. Lauvergeat, S. Hamdi, B. Gallois, Crystal structure of grape dihydroflavonol 4-reductase, a key enzyme in flavonoid biosynthesis, J. Mol. Biol. 368 (2007) 1345–1357.
[135] R.W. Welford, I.J. Clifton, J.J. Turnbull, S.C. Wilson, C.J. Schofield, Structural and mechanistic studies on anthocyanidin synthase catalysed oxidation of flavanone substrates: the effect of C-2 stereochemistry on product selectivity and mechanism, Org. Biomol. Chem. 3 (2005) 3117–3126.
[136] M.L. Lesnick, V.L. Chandler, Activation of the maize anthocyanin gene *a2* is mediated by an element conserved in many anthocyanin promoters, Plant Physiol. 117 (1998) 437–445.
[137] M.L. Falcone Ferreyra, E. Rodriguez, M.I. Casas, G. Labadie, E. Grotewold, P. Casati, Identification of a bifunctional maize C- and O-glucosyltransferase, J. Biol. Chem. 288 (2013) 31678–31688.
[138] M. Jin, X. Zhang, M. Zhao, M. Deng, Y. Du, Y. Zhou, S. Wang, T. Tohge, A.R. Fernie, L. Willmitzer, Y. Brotman, J. Yan, W. Wen, Integrated genomics-based mapping reveals the genetics underlying maize flavonoid biosynthesis, BMC Plant Biol. 17 (2017) 17.
[139] Y. Li, P. Li, Y. Wang, R. Dong, H. Yu, B. Hou, Genome-wide identification and phylogenetic analysis of Family-1 UDP glycosyltransferases in maize (*Zea mays*), Planta 239 (2014) 1265–1279.
[140] J. Nash, K.R. Luehrsen, V. Walbot, Bronze-2 gene of maize: reconstruction of a wild-type allele and analysis of transcription and splicing, Plant Cell 2 (1990) 1039–1049.